"Development of National CRIS Systems in Latin America and the Caribbean: Case Studies 2023"

Rosalina Vázquez Tapia

Autonomous University of San Luis Potosi, Mexico.

PhD student at the University of Salamanca, Spain.

alinavt@usal.es; alinavn@uaslp.mx

May 15-17, 2024

Introduction

- In recent years, the number of CRIS/RIM systems in Latin America has increased, maintaining an upward trend.
- According to the results of a diagnostic survey on the development and characterization of CRIS systems in Latin America [1], a total of 65 CRIS systems were found, and 26 records in the Directory of Research Information System (DRIS) until May 2022.
- Today (May 2024), a total of 39 CRIS systems are registered in the DRIS, which indicates an increase of 50% in two years, the number of registered national CRIS systems doubled, and in the case of institutional ones, there was an increase of 30.64%.
- This paper constitutes the second phase of the diagnostic study that aims to establish a state of the art on the development of institutional and national Management and Research Systems (CRIS) in Latin American and Caribbean countries.

1. Vázquez R. Development and characterisation of CRIS systems in Latin America: Preliminary results of diagnostic survey. Procedia Computer Science. 2022 Jan 1;211(C):267–76.

Particular Objetives

1. Identify the Latin American countries that have a research project or management system or scientific production portal at the national level, and what it consists of, its scope, objectives, development axes, services and benefits.
2. Determine the general characteristics of CRIS projects or systems in operational terms, considering technical, regulatory, methodological and information management aspects.
3. Carry out an analysis, and description of each case according to the established variables and indicators, identifying the similarities or common aspects, and contrasts between the different cases.
4. Determine the implementation processes, results, areas of opportunity, and good practices of each case study, which contribute to a model proposal for the design and implementation of a CRIS system.

Methodology

To identify, analyze and describe national CRIS/RIM systems and similar platforms in Latin American and Caribbean countries, the qualitative descriptive case study method was used, comprising the following activities:

a. Validation of the cases of national CRIS systems and similar projects, which were identified during the first phase of the diagnostic study (2021-2022 Survey), through personal communications with the leaders or managers in each case.

b. Data collection and information analysis by exploring the websites of the national science and technology agencies of the rest of the member countries of LA Referencia and the RedCLARA Network, as well as DRIS and software solutions (Pure, VIVO, DSpace-CRIS, and Sympletic).

Methodology

c. Design of a guide for the collection of information that was sent by e-mail to the people responsible or contact persons of each of the identified projects.

d. Compilation of data, documents, sources of information, and bibliographic references of each case, based on the guide.

e. Documentary and descriptive analysis of each case study, considering the following factors and indicators:

- Project Description: Name, responsible entity, background, objectives, stages of development, lines of work, scope, vision, financing, and strategic alliances.
- Services & Benefits: Participants, adherence procedures, services, benefits, training strategies and resources.
- Regulations and information management: Law and public policy framework, institutional policies and procedures, type of data, and sources of information.
- Technology platform: software, metadata, vocabularies, technical guidelines, and interoperability.

Results

- 7 case studies were determined: 5 National CRIS Systems in Peru, Brazil, Ecuador, Panama, and Puerto Rico; and 2 platforms in Argentina and Chile, respectively, for the management and evaluation of national research and/or scientific production.

No.	Country	Type of system	Name and web site of CRIS/RIM/National Platform	Responsible Agency
1	Argentina	Platform	SIGEVA - Sistema Integral de Gestión y Evaluación. https://sigeva.conicet.gov.ar/	Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET)
2	Brazil	CRIS	BrCRIS – Ecosistema do Pesquisa Brasileira. https://brcris.ibict.br/	Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT)
3	Chile	Platform	DATACIENCIAS – Dimensiones de la Producción Científica Nacional. https://dataciencia.anid.cl/	Agencia Nacional de Investigación y Desarrollo (ANID)
4	Ecuador	CRIS	REDI - Repositorio Semántico de Investigadores del Ecuador. https://redi.cedia.edu.ec/	Corporación Ecuatoriana para el Desarrollo de la Investigación y la Academia (CEDIA)
5	Panama	CRIS	CONECTO – Perfiles de la Ciencia y Tecnología de Panamá. http://conecto.senacyt.gob.pa/conecto/	Secretaria Nacional de Ciencia, Tecnología e Innovación (SENACYT)
6	Peru	CRIS	#PerúCRIS https://perucris.concytec.gob.pe/perucris/presentacion	Consejo Nacional de Ciencia, Tecnología e Innovación Tecnológica (CONCYTEC)
7	Puerto Rico	RIM	BEACON – Sistema de Gestión de Perfiles de Investigadores de Puerto Rico. https://prsciencetrust.org/beaconplatform/	Fideicomiso de Ciencia, Tecnología e Investigación de Puerto Rico

Case study: Peru

#PerúCRIS

Development period: 2012-2022

Objective: To develop an open, interoperable, and integrated national network that provides timely and reliable information services for the benefit of the scientific and academic community and the public and private sectors of Peru.

Information Management: It compiles information on Peru's Science, Technology and Innovation capacities and activities, divided into five main categories:

Human talent, scientific production, institutions, infrastructure, and project.

Services:

- Two types of adhesion procedures, through interoperability with CRIS systems or data loading and editing.
- Information resources, and training activities.

Case study: Peru

#PerúCRIS

Technology Platform:

- #PerúCRIS's technology platform is based on the use of *open-source* software solutions and cloud service architecture.
- It is implemented with DSpace-CRIS software and SuperSet, Dremio (DataLeak), WordPress, and Mule tools.
- It is an interoperable platform that offers Harvest functionalities; exchange of information through the CERIF XML format and the OAI-PMH protocol; validation of external sources, REST API services, SOAP Web services, MULE service bus, and integration with other systems and networks.

Case study: Brazil

The Brazilian Scientific Research Information Ecosystem, BrCris

Development period: 2015 – 2022

Objective: To establish a single model for organizing scientific information for the entire Brazilian research ecosystem, with the aim of providing centralized and up-to-date data on academic, scientific and technological activities of its researchers.

Information Management:

- Actors in the information ecosystem include researchers, projects, infrastructures, laboratories, and research institutions, funders, as well as research results expressed mainly in scientific publications, theses, dissertations, scientific datasets, software, and patents.
- BrCRIS is fed by different sources of information and data, such as: the Lattes Platform (database of CVs of academics and researchers); OasisBr (National Open Access Repository of Publications of Brazil); BDTD (Digital Thesis and Dissertation Database); and other catalogs and directories.

Case study: Brazil

Services:

BrCris offers a unified interface for searching for information, visualization of collaborative networks, and dashboards of science, technology, and innovation indicators.

Technology Platform:

- BrCRIS is primarily implemented on the LA Referencia platform, for data harvesting, deduplication, conversion, storage, and export, using the OAI-PMH protocol.
- The VIVO software is also used as the visualization platform and the VIVO ontology as the basis for the semantic data model.
- It adopts a two-tiered data representation schema, the first being the logical level, a CERIF-based model of entities and relationships, and the second, the *semantic level,* a subset of the VIVO ontology plus a local extension, for data visualization and navigation as a knowledge graph.
- An information search and retrieval interface based on Elastic's Search-UI to create fully customizable interfaces with advanced filtering and sorting of results.

Case study: Ecuador

Ecuador's Semantic Repository for Researchers, REDI

Development period: 2010-2022

Objective: Provide information on authors, publications and areas of knowledge of the Higher Education Institutions that are members and non-members of the Ecuadorian Corporation for the Development of Research and Academia (CEDIA).

Information Management: Information resources on institutions, researchers, publications, projects, datasets, and Patents.

Services:

- REDI makes it possible to extract information from Ecuadorian researchers and their publications, detect similar areas of knowledge, and generate visualization models for the end user.
- It offers search services, statistics, researcher profiles, visualization by research areas, collaboration networks, geospatial information, inter-institutional collaboration by projects, and research trends.

Case study: Ecuador

Ecuador's Semantic Repository for Researchers, REDI

Technology Platform:

REDI is built on a LDP (Linked Data Platform) called Apache Marmotta, which allows the extraction and integration of information, as well as the detection of patterns and the handling of queries, and the implementation of semantic and machine learning technologies.

The platform architecture comprises 5 components:

1) HARVEST: Extraction of data from authors of academic and research papers found in the DSPACE digital repositories of Higher Education Institutions.
2) ENRICHMENT: Extraction of publications through APIs.
3) PROCESSING: Population of an ontology to generate the central graph for the export of information.
4) PERSIST: Detection of similar areas of knowledge, through data mining techniques and algorithms.
5) VISUALIZATION: Visualization models, that allow one to browse and discover researchers, publications, related authors, and areas of knowledge.

Case study: Panama

CONECTO

Development period: The Platform was launched in September 2022.

Objective: CONECTO is a system of scientific and technological profiles of the Republic of Panama, which seeks to encourage collaboration among scientists, students, professors, and the general public. Through this platform, users can create profiles, share information, publish research, and collaborate in joint projects.

Information Management: CONECTO collects different types of information: user profiles, projects, publications, events, financing and employment opportunities, and educational resources.

Services: Indicators of scientific production, metadata curation, alternative metrics, reports (CVU/researcher's resume), training, and statistics.

Technology Platform: CONECTO is implemented with the open-source VIVO software and it uses all the functionalities offered by this solution.

Case study: Puerto Rico

Management System of Researcher Profiles of Puerto Rico, BEACON

Development period: The Platform was launched in October 2021

Objective: BEACON is a centralized platform that collects, displays, analyzes, and reports on all academic activity taking place in Puerto Rico.

Information Management: BEACON automatically harvests publications, and other research output information from online bibliographic databases such as arXiv, CrossRef, ORCID, PubMed, RePEc, Scopus, and Web of Science.

Services: BEACON allows the researcher to add information to their profile manually, export their CV in different formats, locate their publications in different subscribed databases, add publications manually, set up auto-claiming, and manage delegates, preferred name and Altmetric data. To request access, filling out a form is needed.

Technology Platform: The BEACON platform is supported by the Digital Science's SYMPLECTIC software.

Case study: Argentina

Integrated Management and Evaluation System, SIGEVA

Development period: SIGEVA was developed in 2005.

Objective: SIGEVA is a set of computer applications for the management of scholarships, projects, reports, admissions, teaching assessments, institutional memory, and other aspects related to various calls.

Information Management: The system comprises a set of information modules that interact following a data flow and are interoperable with the systems of member universities and institutions. It is interoperable with the Cvar platform for the management of curricula vitae of researchers.

Services: They are mostly offered via Intranet to member institutions, according to their requirements. The form of adhesion is through an institutional agreement with the National Council for Scientific and Technical Research (CONICET).

Technology Platform: The software solution is developed in-house, using metadata standards, persistent digital identifiers, and interoperability guidelines.

Case study: Chile

DataCiencia

DATACIENCIA
Dimensiones de la Producción Científica Nacional

Development period: The platform was developed in 2014

Objective: The DataCiencia platform aims to represent, within its multiple dimensions, the Chilean scientific production, according to publications indexed since 2008 on WoS, Scopus, and the SciELO Chile collection. Its main goal is to visualize and facilitate indicators of national scientific production free of charge.

Information Management:

Dataciencia obtains most of its data (scientific publications) from three international databases: WoS, Scopus, and SciELO-Colección Chile. The managed data correspond to scientific articles and metrics from the profiles of researchers and institutions.

Services:

It provides up-to-date metrics of scientific production in Chile; it allows to generate reports in real time; it has an author disambiguation feature; it offers detailed information on production and citations by region, as well as regional profiles and scientific collaboration networks based on co-authorship.

Case study: Chile

DataCiencia

DATACIENCIA
Dimensiones de la Producción Científica Nacional

Technology Platform:

- It is self-developed.
- It is carried out using Ruby on Rails for the front end, and Flask for the back end. For data warehousing, two different databases are used, Elasticsearch, and MySQL. Apache web server and API development.
- It uses persistent digital identifiers: DOI and WOS/Scopus, and PID SciELO-Chile identifiers for articles; ISSN for journal profiles, and the integration of ORCID into author profiles.

Results

- The three countries with the highest number of CRIS systems are Peru with one national system, and 24 institutional systems; Chile with 12 institutions, and a national platform; and Mexico with 11 institutions.
- The most widely used software solution by countries in the region is PURE, with a total of 36 installations, followed by VIVO with 17; other solutions with 14, DSpace-CRIS with 12, and Sympletic with 2.
- The national CRIS systems with the greatest scope and development are #PerúCRIS from Peru, and BrCRIS from Brazil. Both are implemented with open-source platforms, and open and interoperable technologies with their repositories, databases, journal portals, and external sources.
- Except for Puerto Rico, the rest of the countries with national CRIS systems use open-source software solutions (DSpace-CRIS and VIVO), open technologies, and the CERIF standard as part of their architecture.

Conclusions

- In all cases, each country has developed its own model, in an average time of 2 to 8 years, with different scopes and approaches, but with the common objective of integrating, managing, and giving visibility to the scientific production of their researchers and institutions, providing them with a set of services, indicators, and metrics.

- As a future work, in order to establish a state of the art on the development of Management and Research Systems (CRIS) at an institutional and national level, in Latin American and Caribbean countries, a follow-up will be necessary with each institution/country based on the results of the 2022 diagnostic survey (first phase of the study), and the present case study (second phase), considering the following research activities:

  1. Verification of the 65 CRIS systems identified in the 2022 study, to validate those that are operating and those that could not be found or are inactive.
  2. Follow up on the institutions that responded in the 2022 survey that did not have a CRIS system, but that did have their development planned in the medium or long term.

Conclusions

3. Repeat searches on the websites of DRIS, LYRASIS (VIVO), 4SCIENCE (DSpace-CRIS), and ELSEVIER (PURE), to identify new registered CRIS systems and installations of these software solutions.
4. Expand the selection criteria, adding other software installations designed or adapted for CRIS systems, considering both commercial and open-source solutions.
5. Validate and complement the information documented in each case study, through interviews with the responsible personnel and technical staff of each of the seven national systems/platforms.

16th International Conference on Current Research Information Systems (CRIS2024)
Vienna, Austria; May 15-17, 2024

Thank you!

Rosalina Vázquez Tapia

Researcher at the Autonomous University of San Luis Potosi, Mexico.
General Coordinator of the Mexican Network of Institutional Repositories - REMERI
Technical Responsible of the Mexican node in the Federated Network of Institutional Repositories of Scientific Publications - LA Referencia
PhD student at the University of Salamanca, Spain.
alinavt@usal.es; alinavn@uaslp.mx