



Artículo científico

Referencia

Lajpop Ajpacaja, K. A., Ixcolin Ramírez, A. E. y Ortiz Álvarez, R. O. (2024). Arquitectura de software de un sistema de investigación actual para la USAC *Revista Científica del Sistema de Estudios de Postgrado*. 7(1). 15-25. DOI: <https://doi.org/10.36958/sep.v7i1.210>

Arquitectura de software de un sistema de investigación actual para la USAC

Software architecture of a current research system for USAC

Kevin Adiel Lajpop Ajpacajá

Maestro en Tecnologías de la Información
y Comunicación
Universidad de San Carlos de Guatemala
kevinadiel@gmail.com
<https://orcid.org/0000-0002-0392-7632> 

Ana Elizabeth Ixcolin Ramírez

Licenciada en Administración y Gerencia de
Servicios de Salud
Universidad Galileo
annyram3@gmail.com
<https://orcid.org/0009-0007-4658-0468> 

Rony Ormandy Ortiz Álvarez

Estudiante de Ingeniería en Ciencias y Sistemas
Universidad de San Carlos de Guatemala
ormandyrony@gmail.com
<https://orcid.org/0009-0008-4851-0161> 

Recibido: 01/09/2023

Aceptado: 20/05/2024

Publicado: 24/06/2024

Resumen

OBJETIVO: generar una arquitectura de software que soporte las necesidades de un sistema de información que gestiona la investigación actual (CRIS) en la Universidad de San Carlos de Guatemala (USAC). **MÉTODO:** se tomaron requerimientos con base en reuniones con actores importantes de la ciencia en la USAC, esto con el fin de poder obtener sus necesidades y compararlas con las características de software que presentan las distintas soluciones de software libre de CRIS. Posterior a ello, con requerimientos claros y características cumplidas se construyó la arquitectura de los sistemas de información que son fuentes de datos e información primaria para el CRIS. **RESULTADOS:** como principal producto se tuvo el diseño arquitectónico de todos los sistemas de información que forman parte de un CRIS para la USAC, dicha arquitectura cumple las necesidades que tiene la universidad para tener un sistema de este tipo que gestiona la investigación. **CONCLUSIÓN:** el primer paso para la adopción e implementación de un sistema CRIS es la investigación de las necesidades de la institución para posteriormente compararlas con las soluciones CRIS para diseñar una arquitectura, es importante también tener en cuenta que no todos los diseños arquitectónicos son ideales para todas las instituciones, al contrario, cada una de ellas debe de contar con un diseño de sistema de información único.

Palabras clave

CRIS, ciencia, sistemas de Información, USAC

Abstract

OBJECTIVE: generate a software architecture that supports the needs of an information system managing current research (CRIS) at the University of San Carlos of Guatemala (USAC). **METHOD:** requirements were gathered through meetings with key stakeholders in science at USAC, aiming to understand their needs and compare them with the software features of different CRIS open-source solutions. Following this, with clear requirements and met characteristics, the architecture of information systems serving as primary data sources for CRIS was built. **RESULTS:** the main outcome was the architectural design of all information systems forming part of a CRIS for USAC, fulfilling the university's needs for managing research. **CONCLUSION:** the initial step towards adopting and implementing a CRIS system is researching the institution's needs, then comparing them with CRIS solutions to design an architecture. It's also crucial to note that not all architectural designs are suitable for every institution; each should have a unique information system design.

Keywords

CRIS, science, information systems, USAC

Introducción

La investigación es una pieza importante para el desarrollo de un país y por ende la buena gestión de esta es vital, por lo que un sistema de información de investigación actual (CRIS) es una herramienta vital para lograr los objetivos de la ciencia. Esta herramienta o tipo de tecnología no es nueva, de hecho, tiene más de 40 años de estarse utilizando, pues siempre ha sido importante tener el control de una investigación desde que empieza su formulación hasta su divulgación.

La implementación de un CRIS debe de iniciar con el uso de las buenas prácticas, que indican plantear una ruta a seguir para el desarrollo de un CRIS, lo cual debe de contener el alcance que tendrá, modelo de datos y la arquitectura de software.

Esto se plantea bajo los estándares y módulos mínimos que se han definido por instituciones que ya han implementado un CRIS.

Para lograr esta investigación, es necesario observar cómo se comporta la USAC actualmente; después, plasmar ese comportamiento en procesos y un modelo arquitectónico que represente el desarrollo de la investigación.

Guatemala no cuenta con indicadores a nivel internacional de ciencia, estadísticas reales sobre las investigaciones y sobre todo resultados que tienen dichas investigaciones, tampoco forma parte del conjunto de países que cuenta con un CRIS y mucho menos forma parte de las estadísticas de cuál es el porcentaje de cumplimiento o de objetivos cumplidos por medio de un CRIS, esto hace que la iniciativa de la USAC de contar con un CRIS sea pionera y estandarte a nivel nacional de la ciencia e investigación.

Materiales y métodos

Antes de definir una arquitectura para un CRIS se debe de definir con qué componentes debe contar una implementación de CRIS para ser denominado un sistema que gestione la investigación actual.

Para ello se siguió la definición de Dvořák, J., et al., (2019) quién dicta doce partes esenciales que un CRIS debe de cubrir, siendo estas: bases de datos de artículos científicos, métricas de citas de los artículos y revistas, repositorio institucional de acceso abierto, vinculación con ORCID, sistemas de gestión de identidad única, sistemas de gestión de recursos humanos, sistemas de gestión y vinculación de estudiantes, sistemas para la gestión de finanzas, gestión documental, sistemas para la gestión de financiamiento interno, oficina de transferencia tecnológica y manejo de financiamiento externo. Es necesario priorizar para las necesidades que presenta la USAC de la siguiente manera:

Tabla 1

Necesidades y características de un CRIS

1. Sistemas para la gestión de financiamiento interno.	7. Métricas de citas de los artículos y revistas.
2. Gestión documental.	8. Repositorio institucional de acceso abierto
3. Vinculación con ORCID.	9. Manejo de financiamiento externo.
4. Sistemas de gestión y vinculación de estudiantes	10. Sistemas para la gestión de finanzas.
5. Sistemas de gestión de identidad única o SSO.	11. Sistemas de gestión de recursos humanos.
6. Bases de datos de artículos científicos	12. Oficina de transferencia tecnológica.

Es necesario tener en cuenta que el orden de prioridades se obtuvo partiendo de las necesidades de la universidad, pero también la madurez tecnológica de la misma, por ejemplo, una oficina de transferencia tecnológica por el momento supera las capacidades administrativas de la universidad o un sistema de gestión de recursos humanos, no es del todo provechoso para el nivel actual de madurez tecnológica de la universidad.

Contar con una calidad del dato en un CRIS no solo es una buena práctica sino también apoya a la generación de nueva investigación, contar con un CRIS con datos de alta calidad genera más investigación, como lo es el estudio de Vaci, N., et al. (2020), la cual está centrada en procesar el lenguaje natural, pero con el extra de que estos datos vienen del UK-CRIS, ejemplificando que en un CRIS consolidado no solo es un repositorio para la comunidad científica, sino también es una fuente de datos e investigación, Vaci, N., et al. (2020).

Asserson, A., et al. (2002) define que los datos deben de ser intercambiables y accedidos, bajo esta premisa Azeroual, O. (2019) define algunos fundamentos que los datos de un CRIS deben cumplir como las afirmaciones:

- a) Los datos son de calidad cuando son aptos para que puedan ser utilizados.
- b) También deben de estar en el lugar, momento y formato correcto para poder ser utilizados.
- c) Bajo estas premisas los datos deben de cumplir con las siguientes características: integros, correctos, consistencia, oportunos.

Las necesidades obtenidas en función de lo que menciona Dvořák, J., et al., (2019), deben de ser comparadas contra las funciones que presenta una solución DSpace-CRIS, en función de ello se crea la arquitectura que servirá de cimiento para la construcción e implementación de un sistema CRIS para la Universidad de San Carlos de Guatemala.

Resultados y discusión

Se realizaron varias jornadas de entrevistas con el fin de obtener información, requerimientos y necesidades de los distintos centros de investigación los cuales se resumen en los siguientes:

- Buscar actualización de los investigadores
- Investigación compartida
- Descentralización en la universidad
- Apoyo de la editorial universitaria
- Registros por centros o por departamentos de centros de investigación.
- Es necesario tener esta plataforma, para sentirse parte de una red local, regional, nacional o internacional
- Migración de informes de la DIGI al nuevo sistema
- Capacidad para las evaluaciones por pares, como se puede reforzar la fase de evaluación por pares
- Gestión de usuarios o jerarquía de usuarios como un tipo de administrador para gestionar la información.

Varios requerimientos se mencionaron en más de una ocasión, por ejemplo, la descentralización y no depender de una administración central para el manejo de los datos, también se hace mucho énfasis en la visibilidad de las investigaciones y producción científica que realiza la universidad, esto empata con lo que comenta Vázquez, R. (2022), al momento de no contar con un sistema CRIS, se corre el riesgo de aislar la investigación que se realiza en un país o universidad.

Los requisitos identificados por los diversos actores principales deben ser abordados por una solución de CRIS. Para ello, se examinan las cualidades de la solución DSpace-CRIS junto con los requisitos recolectados. En el cuadro siguiente, se efectúa esa comparación entre los requisitos y características que se derivan de la documentación oficial de DSpace-CRIS (Lyrisis, 2023).

Tabla 2

Requerimientos y necesidades principales

Característica	Requerimiento	DSpace-CRIS
Propósito	Buscar actualización de los investigadores	Gestión de información de investigación y repositorio
Tipos de Contenido	Migración de informes en la DIGI al nuevo sistema	Resultados de investigación, proyectos, financiamiento, publicaciones, CV del investigador, la información de gestión, los informes a los financiadores, la bibliografía de investigación y los informes de resultados comerciales
	Registrados por centros o por departamentos de centros de investigación	

Soporte de Metadatos	Apoyo de la editorial universitaria	Metadatos mejorados para información de investigación
Integración	Buscar actualización de los investigadores Investigación compartida Registrados por centros o por departamentos de centros de investigación.	ORCID, DOI, identificadores de investigación, sistemas externos
Informes	Es necesario tener esta plataforma, para sentirse parte de una red local, regional, nacional o internacional	Análisis avanzados para actividades e impacto de investigación
Perfiles de Investigadores	Buscar actualización de los investigadores Es necesario tener esta plataforma, para sentirse parte de una red local, regional, nacional o internacional Capacidad para las evaluaciones por pares, como se puede reforzar la fase de evaluación por pares	Perfiles de investigadores y funciones de redes
Personalización		Modelos de datos personalizables para información de investigación
Integración de Flujo de Trabajo		Integración con flujos de trabajo institucionales y CRIS
Controles de Acceso	Gestión de usuarios o jerarquía de usuarios como un tipo de administrador para gestionar la información	Controles de acceso similares a DSpace
Extensibilidad		Altamente extensible para adaptarse a diversas necesidades
Casos de Uso	Investigación compartida Buscar actualización de los investigadores Registrados por centros o por departamentos de centros de investigación	Gestión integral de investigación y presentación
Estándares	Migración de informes en la DIGI al nuevo sistema. Apoyo de la editorial universitaria	DSpace-CRIS cumple con el estándar CERIF

Se puede observar que existe una correlación entre las necesidades expuestas y las características principales que presenta DSpace-CRIS y, de hecho, existen características que no han sido exploradas en su totalidad como es el caso de la extensibilidad, la personalización o los flujos de trabajo que maneja DSpace-CRIS, esto da una oportunidad para se pueda seguir explorando.

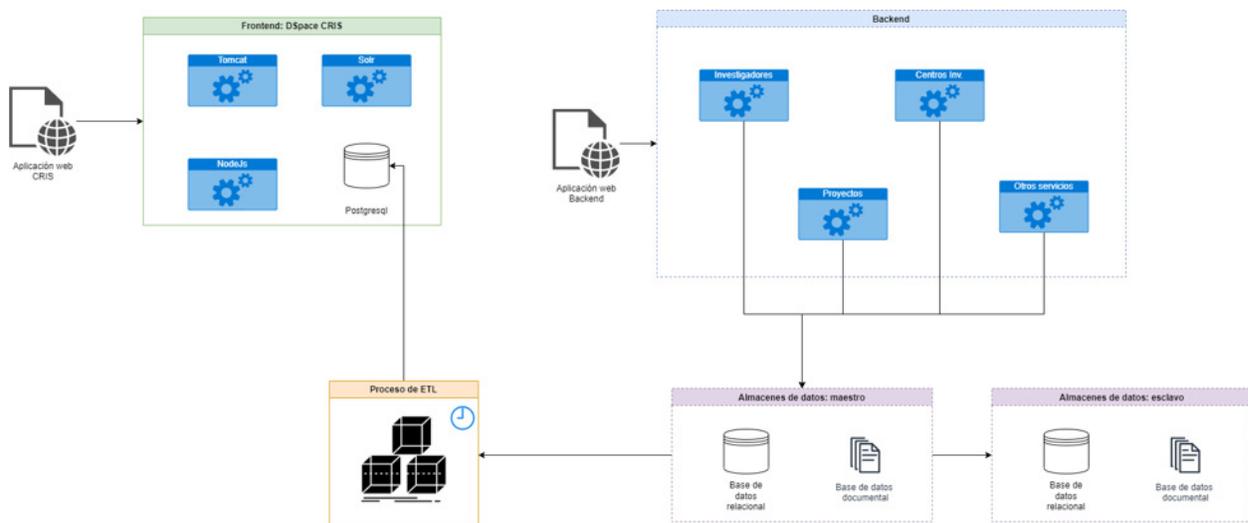
Arquitectura de software

La arquitectura de software está compuesta por dos grandes escenarios, la parte del frontend y la del backend, este último es el encargado de gestionar y generar toda la información transaccional y documental que modelan los procesos digitales de un sistema de investigación (Figura 1).

Mientras que la parte del frontend es el CRIS, en este caso se toma en cuenta la solución open source DSpace-CRIS el cuál sigue un estándar CERIF para datos y una implementación que se encarga de gestionar, administrar y sobre todo mostrar información científica.

Figura 1

Arquitectura de software para un CRIS



Backend

El backend es la fuente primera de estos, como lo es el sistema de investigadores, sistema de gestión de proyectos, sistema de gestión de centros de investigación.

Es necesario que estos servicios existan, dado que la naturaleza de un proyecto de investigación radica en los investigadores y centros de investigación, siguiendo la línea de Dvořák, J., et al., (2019).

La arquitectura desarrollada para la parte del backend es un patrón arquitectónico de microservicios que pueden ser más granulares, dependiendo de la solución que se implemente. Entre los microservicios mínimos para que se gestione la investigación vía un sistema de información se tienen:

Servicio de investigadores: encargado de gestionar a los investigadores, teniendo en cuenta funciones como actualización de datos personales, académicos o demográficos, en otras palabras, toda la ficha personal y científica del investigador.

Servicio de proyectos: encargado de la gestión de proyectos de investigación que reciben financiamiento concursable, en este apartado del sistema de información se debe de gestionar todo el proceso de evaluación hasta su aprobación del proyecto, como mínimo, para que genere datos para un sistema CRIS.

Servicio de centros de investigación: encargado de la gestión de estos, en el cual aplica altas, bajas y cambios de los datos del centro de investigación, este es otra fuente primaria de datos, también es una parte esencial para los proyectos.

Otros servicios: Otros servicios que puedan ser necesarios o complementarios, como por ejemplo el servicio de reportes, servicios que gestionen notificaciones u otros.

El backend también utilizará componentes de almacenamiento que incluyen:

Almacén de datos maestro: todos los datos que se generen en los servicios descritos anteriormente se guardarán en dos almacenes de datos, una base de datos relacional que tendrá todos los datos de tipo transaccionales y una base de datos documental que tendrá todos los datos que son de tipo no estructurado o documental.

Almacén de datos esclavo: es una réplica exacta en línea tanto de base de datos relacional como base de datos documental, lo que se busca es el respaldo en tiempo real de los datos y todo lo documental.

Proceso de ETL

En la arquitectura que se presenta, hay un punto clave, la fuente de datos que está en el propio esquema del backend. Estos datos son primarios o crudos que respetan un esquema diferente al necesario para DSpace CRIS, para ello debe de existir un proceso de transformación de estos datos para que sean verdaderamente la fuente de información del sistema CRIS.

Dicho proceso es el encargado de obtener los datos del backend y posteriormente transformarlos y cargarlos al almacén de datos del sistema DSpace-CRIS.

Este proceso se realizará de manera periódica prudencial, porque los datos de investigadores, proyectos y centros de investigación crecerán de manera continua.

Frontend: Dspace CRIS

Al componente CRIS se le denomina frontend, en función que es el componente que estará a la vista de cualquier interesado en la información científica que se publique, bajo esta línea de pensamiento, el software propuesto es DSpace-CRIS.

Hasta el momento es necesario tener en cuenta que no existe una imagen de docker estable (Illig, S., et al., 2017), dado esta situación, la implementación de DSpace-CRIS se recomienda que sea en una implementación monolítica por lo estable a la hora de la implementación.

La fuente de datos de los proyectos la alimentará el proceso de ETL que tomará como datos crudos o primarios los datos generados en los sistemas del backend.

Conclusiones

La arquitectura de un CRIS debe de responder a una serie de requerimientos brindados por los diferentes centros de investigación, porque las necesidades de un sistema CRIS son muy diferentes a las necesidades de un sistema de información convencional, por lo que la toma de requerimientos debe de ir desde la creación del modelo arquitectónico.

Es necesario que exista todo un ambiente digital detrás de un sistema CRIS, no únicamente un CRIS como fuente de información de cara al público y la comunidad académica, sino también tener sistemas de información que permitan la automatización de los procesos de la investigación.

La arquitectura que debe de seguir un CRIS debe de ser altamente escalable y desacoplada que permita al mayor cambio, el menor impacto posible.

Es de vital importancia que el sistema CRIS se alimente constantemente y de manera automática de la fuente de datos primaria, para ello es necesario contar con un proceso de ETL que permita la carga automática.

Agradecimientos

A la Dirección General de Investigación (DIGI) por el financiamiento dado para la realización de este proyecto correspondiente a la convocatoria diferenciada 2023.

Referencias

Asserson, A., Jeffery, K. G., & Lopatenko, A. (2002). CERIF: past, present and future: an overview. <http://hdl.handle.net/11366/131>

Azeroual, O. (2019). Text and data quality mining in CRIS. *Information*, 10(12), 374. <https://doi.org/10.3390/info10120374>

Dvořák, J., Chudlarský, T., & Špaček, J. (2019). Practical CRIS interoperability. *Procedia computer science*, 146, 256-264. <https://doi.org/10.1016/j.procs.2019.01.077>

Illig, S., & Eiermann, A. M. (2017). Running DSpace: DSpace-CRIS with Docker.

Lyrisis. (2023). Introduction to DSpace-CRIS. DSpace-CRIS Wiki. <https://wiki.lyrisis.org/display/DSPACECRIS/Introduction>

Vaci, N., Liu, Q., Kormilitzin, A., De Crescenzo, F., Kurtulmus, A., Harvey, J., ... & Nevado-Holgado, A. (2020). Natural language processing for structuring clinical text data on depression using UK-CRIS. *BMJ Ment Health*, 23(1), 21-26. <http://dx.doi.org/10.1136/ebmental-2019-300134>

Vázquez, R. (2022). Development and characterisation of CRIS systems in Latin America: Preliminary results of diagnostic survey. *Procedia Computer Science*, 211, 267-276. <https://doi.org/10.1016/j.procs.2022.10.201>

Sobre los autores

Kevin Adiel Lajpop Ajpacajá

Tiene un Doctorado en Ciencias de la Computación, Doctorado en Investigación con Énfasis en Tecnología, Máster Tecnologías de la Información y la Comunicación, Especialista en Administración Tecnológica y Ingeniero en Ciencias y Sistemas.

Ana Elizabeth Ixcolin Ramírez

Tiene una Maestría en Salud y Seguridad del Trabajo, Licenciada en Administración y Gerencia en Servicios de Salud, y es Técnico en Visita Médica y Mercadeo.

Rony Ormandy Ortíz Álvarez

Estudiante de Ingeniería en Ciencias y Sistemas Noveno Semestre e Ingeniería Electrónica Séptimo Semestre. Colaboró como Auxiliar de Investigación en la Universidad de San Carlos de Guatemala en el proyecto de diseño e implementación de un sistema CRIS para la

Dirección General de Investigación. Fue asistente de laboratorio del curso de introducción a la programación de computadoras de la escuela de Ingeniería Mecánica Eléctrica. Tercer lugar en el BootCamp Enactus - Pantaleon 2021. Finalista del Rally de innovación interdepartamental para mi comunidad 4.0 de Senacyt.

Financiamiento de la investigación

Financiamiento por la Convocatoria Diferenciada 2023 de la Dirección General de Investigación.

Declaración de intereses

Declara no tener ningún conflicto de intereses, que puedan haber influido en los resultados obtenidos o las interpretaciones propuestas.

Declaración de consentimiento informado

El estudio se realizó respetando el Código de ética y buenas prácticas editoriales de publicación.

Derecho de uso

Copyright (c) (2024) Kevin Adiel Lajpop Ajpacajá, Ana Elizabeth Ixcolin Ramírez & Rony Ormandy Ortiz Álvarez

Este texto está protegido por la [Licencia Creative Commons Atribución 4.0 Internacional](https://creativecommons.org/licenses/by/4.0/)



Este texto está protegido por una licencia [Creative Commons 4.0](https://creativecommons.org/licenses/by/4.0/).

Es libre para compartir, copiar y redistribuir el material en cualquier medio o formato y adaptar el documento, remezclar, transformar y crear a partir del material para cualquier propósito, incluso comercialmente, siempre que cumpla la condición de atribución: debe reconocer el crédito de una obra de manera adecuada, proporcionar un enlace a la licencia, e indicar si se han realizado cambios. Puede hacerlo en cualquier forma razonable, pero no de forma tal que sugiera que tiene el apoyo del licenciante o lo recibe por el uso que hace.