

# Repositorio de datos y sistemas de información para la gestión de la investigación: proyecto piloto en un ámbito universitario



Universidad  
del Cauca

Adrián Felipe Burbano Narvárez  
Guido Naim Imbachi Alvarado

Directora: PhD. Sandra Milena Roa Martínez

**Universidad del Cauca**  
**Facultad de Ingeniería Electrónica y Telecomunicaciones**  
**Departamento de Sistemas**  
**Grupo de Investigación en Inteligencia Computacional (GICO)**  
Línea de investigación: Información y Tecnología  
Enero de 2024

## **AGRADECIMIENTOS**

Inicialmente, expresamos nuestro agradecimiento a Dios por todas las oportunidades que nos ha concedido, por nuestra familia y las personas que ha puesto en nuestras vidas para ayudarnos a crecer y culminar este proceso.

Quisiéramos agradecer a las personas que de alguna forma contribuyeron al desarrollo y la finalización del presente trabajo, de manera especial, queremos agradecer a nuestra directora de tesis PhD. Sandra Milena Roa Martínez, por su orientación, motivación y compromiso con nosotros y nuestro trabajo, su paciencia, aportes significativos, las sugerencias constructivas, el tacto al comunicarnos las observaciones y su dirección han sido elementos fundamentales para realización y culminación de nuestro trabajo. Siempre estuvo dispuesta a ayudarnos, guiándonos con su experiencia, sabiduría y brindando su apoyo incondicional.

Adrián Felipe Burbano – Guido Naim Imbachí

Quiero presentar mis más nobles y sinceros agradecimientos, primero a mi madre Luz Mary Narvárez quien es mi más grande apoyo en todo sentido con su amor incondicional, a mi padre Fabio Burbano que a pesar de estar en el cielo construyó lo necesario para apoyarme en un futuro y que pudiese lograr lo que hasta ahora he alcanzado, a Sebastián Burbano de quien valoro su amistad incondicional como hermano y admiro como persona, también quiero agradecer a Sofía Trujillo, persona que fomento en mí el espíritu de disciplina y constancia académica, y motivo en mí el esfuerzo para conseguir lo que deseo.

Seguidamente a mi compañero de trabajo de grado y a quién considero como mi amigo, Guido Imbachí pues su compañía, apoyo y compromiso fueron imprescindibles en la realización de este proyecto de investigación para completar uno de nuestros sueños el cual es optar por el título de Ingenieros de sistemas. También quiero agradecer a los compañeros que me acompañaron durante mi proceso académico y de los cuales aprendí cada cosa que consideré necesaria tomar tanto en mi vida personal como profesional.

Adrián Felipe Burbano

Quiero extender mi agradecimiento a cada uno de mis queridos hermanos y padres, quienes fueron mis faros en todo momento. Agradecer a las personas que en silencio me respaldaron y motivaron impulsándome a superar los desafíos desde el primer día de mi ingreso a mi formación de pregrado. Así mismo, manifestar mi gratitud a la respetable familia Muñoz Cantillo, a mis profesores y compañeros de formación.

De igual manera, mi reconocimiento a mi compañero y amigo Adrián Burbano, cuya dedicación y empeño han sido un baluarte, un constante ejemplo y motivo de admiración. Su colaboración y capacidad para encontrar soluciones en momentos críticos fueron fundamentales para superar obstáculos y alcanzar metas a lo largo del desarrollo de este trabajo.

Finalmente, agradezco a todos aquellos que, de alguna manera, han contribuido a mi formación y han sido parte esencial de mi viaje académico. El llegar hasta aquí no es un logro personal, sino de cada persona que ha iluminado mi camino.

Guido Naim Imbachí

## TABLA DE CONTENIDO

1. INTRODUCCIÓN .....	1
1.1 Justificación.....	3
1.2 Objetivos.....	3
1.3 Aportes.....	4
2. DE UN MAPEO SISTEMÁTICO A UNA REVISIÓN SISTEMÁTICA DE LA LITERATURA .....	5
2.1 Procedimiento metodológico.....	5
2.1.1 Cadena de búsqueda.....	5
2.1.2 Criterios de inclusión y exclusión .....	6
2.1.3 Criterios para la evaluación de la calidad .....	6
2.2 Resultados .....	9
2.2.1 Artículos por año.....	9
2.2.2 Licencias y derechos de autor .....	9
2.2.3 Estándares de metadatos, normas y modelos.....	11
2.2.4 Repositorios de datos de investigación.....	12
2.2.5 Current Research Information System (CRIS).....	15
2.3 Hacia una revisión sistemática .....	16
2.3.1 Datos de investigación .....	16
2.3.2 Metadatos.....	25
2.3.3 Estándares de metadatos.....	27
2.3.4 Repositorios de datos de investigación.....	29
2.3.5 Sistemas de información de investigación actual (CRIS).....	35
2.3.6 Consideraciones finales de la revisión sistemática.....	37
3. PROPUESTA DE MODELO DE PUBLICACIÓN DE DATOS DE INVESTIGACIÓN PARA UN ÁMBITO UNIVERSITARIO .....	40
3.1 Modelos de representación de datos.....	40
4. PROTOTIPO DE REPOSITORIO DE DATOS DE INVESTIGACIÓN PARA LA UNIVERSIDAD DEL CAUCA INVESTUNICAUCA.....	54
4.1 Marco de trabajo para el desarrollo del prototipo .....	54
4.1.1 <i>Backlog</i> del Producto.....	54
4.1.2 Definición de Sprint.....	55

4.2	Definición del prototipo de repositorio de datos de investigación INVESTUNICAUCA.....	56
4.3	Perfil de aplicación del prototipo INVESTUNICAUCA.....	56
4.4	Aspectos tecnológicos.....	57
4.4.1	Software.....	58
4.4.2	Definición de formatos de archivos el prototipo INVESTUNICAUCA.....	59
4.5	Implementación del Prototipo.....	60
4.5.1	Precondiciones.....	60
4.5.2	Instalación.....	60
4.5.3	Persistencia.....	61
4.5.4	Instalación <i>Backend</i> .....	62
4.5.5	Instalación <i>Frontend</i> .....	65
4.6	Configuración del repositorio para un grupo de investigación. ....	66
4.6.1	Instanciación del perfil de aplicación en el prototipo INVESTUNICAUCA.....	67
4.6.2	Definición de identificadores persistentes para el prototipo del repositorio de datos de investigación INVESTUNICAUCA.....	71
4.6.3	Definición de entidades CERIF para el prototipo de repositorio de datos de investigación INVESTUNICAUCA.....	73
4.6.4	Definición y estandarización de formatos de archivos para el prototipo del repositorio de datos de investigación INVESTUNICAUCA.....	75
4.6.5	Definición de licencias para el prototipo del repositorio de datos de investigación INVESTUNICAUCA.....	76
4.6.6	Personalización de interfaz de usuario del prototipo del repositorio de datos de investigación INVESTUNICAUCA.....	77
4.7	Revisión de funcionalidades de gestión de un ítem en el prototipo INVESTUNICAUCA.....	85
4.7.1	Definición de casos diseñados para la revisión de software. ....	85
5.	CONCLUSIONES.....	90
6.	TRABAJOS FUTUROS.....	92
7.	REFERENCIAS.....	93
8.	ANEXOS.....	105

## ÍNDICE FIGURAS

Figura 1. Número de artículos por año.....	9
Figura 2. Cantidad de artículos por cada licencia de derechos de autor. ....	10
Figura 3. Cantidad de publicaciones que abordan estándares de metadatos. ....	12
Figura 4. Cantidad de publicaciones que abordan repositorios de datos de investigación.....	14
Figura 5. Cantidad de publicaciones que abordan plataforma para creación de repositorios de datos de investigación. ....	15
Figura 6. Cantidad de publicaciones por cada plataforma CRIS.....	16
Figura 7. Modelo de datos de Dspace. Fuente: Tomado de: (LYRISIS, 2011) ..	45
Figura 8. Modelo Conceptual de DSpace- CRIS para INVESTCAUCA.....	47
Figura 9. Modelo lógico DSpace-CRIS para INVESTCAUCA.....	52
Figura 10. Interfaz gráfica de la plataforma de búsqueda Solr para el prototipo INVESTUNICAUCA.....	64
Figura 11. Interfaz gráfica del servidor web Tomcat para el prototipo INVESTUNICAUCA.....	64
Figura 12. Captura de pantalla de la interfaz gráfica de la instalación dspace-cris para prototipo INVESTUNICAUCA.....	66
Figura 13. Captura de pantalla de la interfaz gráfica para definición de metadatos del prototipo INVESTUNICAUCA.....	68
Figura 14. Captura de pantalla de la interfaz gráfica con la definición de metadatos datacite del prototipo investunicauca.....	69
Figura 15. Captura de pantalla de la interfaz gráfica metadato agregado a datacite de prototipo investunicauca. ....	70
Figura 16. Captura de pantalla de la interfaz gráfica colección con nuevo metadato de prototipo INVESTUNICAUCA.....	71
Figura 17. Captura de pantalla de la interfaz gráfica metadatos de identificadores persistentes en la creación de un ítem del prototipo INVESTUNICAUCA.....	72
Figura 18. Captura de pantalla de la interfaz gráfica opciones de identificadores persistentes en la creación de un ítem del prototipo INVESTUNICAUCA.....	72
Figura 19. Captura de pantalla de archivos para definición de entidades CERIF en prototipo INVESTUNICAUCA.....	74
Figura 20. Captura de pantalla de formatos de archivos en prototipo INVESTUNICAUCA.....	75

Figura 21. Captura de pantalla de archivos para configuración de licencias creative Commons en prototipo INVESTUNICAUCA.....	76
Figura 22. Captura de pantalla de licencia de depósito por defecto en prototipo INVESTUNICAUCA.....	77
Figura 23. Captura de pantalla de código angular personalizado en la interfaz de usuario del prototipo INVESTUNICAUCA.....	78
Figura 24. Personalización de interfaz de usuario del prototipo INVESTUNICAUCA.....	79
Figura 25. Captura de pantalla de configuración servidor SMTP INVESTUNICAUCA.....	80
Figura 26. Captura de pantalla de creación de nuevo usuario y recuperación de contraseña en el prototipo INVESTUNICAUCA.....	81
Figura 27. Captura de pantalla de interfaz gráfica para grupos y roles en el prototipo INVESTUNICAUCA.....	82
Figura 28. Captura de pantalla de interfaz gráfica para asignar roles a una colección en el prototipo INVESTUNICAUCA.....	84

## ÍNDICE TABLAS

Tabla 1.	Documentos obtenidos al aplicar criterios para la evaluación de la calidad a la cadena de búsqueda en la base de datos especializada Scopus.....	7
Tabla 2.	Nombres de licencias de derechos de autor. ....	10
Tabla 3.	Nombre de estándares de metadatos y total de publicaciones. ....	11
Tabla 4.	Repositorios mayormente descritos en revisión bibliográfica.....	13
Tabla 5.	Plataformas para crear repositorios de datos de investigación.....	14
Tabla 6.	Definición de entidades modelo CERIF .....	40
Tabla 7.	Entidades que contribuyen a alcanzar los principios FAIR.....	43
Tabla 8.	Atributos para la entidad persona .....	48
Tabla 9.	Atributos para la entidad Proyecto .....	49
Tabla 10.	Atributos para la entidad Organización .....	49
Tabla 11.	Atributos para la entidad Evento .....	50
Tabla 12.	Atributos para la entidad Patrocinio .....	50
Tabla 13.	Atributos para la entidad Producto.....	50
Tabla 14.	Planificación de Sprints y Tareas .....	55
Tabla 15.	Perfil de aplicación para el prototipo INVESTUNICAUCA .....	57
Tabla 16.	Formatos que soporta la plataforma software y que son requeridos para prototipo INVESTUNICAUCA.....	59
Tabla 17.	Herramientas software para realizar la instanciación de DSpace-CRIS 60	
Tabla 18.	Revisión de funcionalidad agregar ítem en colección.....	86
Tabla 19.	Revisión de funcionalidad eliminar ítem en colección .....	87
Tabla 20.	Revisión de funcionalidad editar ítem en colección .....	88



## ABSTRACT

During the development of scientific research processes, large volumes of data are generated that describe the context of the investigations and support investigative productions. These data can be reused, becoming very useful for the generation of new intellectual productions, for example. Therefore, it is highly relevant information and the implementation of a software solution that helps automate the management of this particular data is necessary. A software solution such as a research data repository provides answers to the needs raised. For this reason, the objective of this work is to develop a prototype research data repository for the publication of data sets linked to elements of current research information systems (CRIS - Current Research Information System) for management of research groups from the University of Cauca. A theoretical foundation is carried out on research data, metadata standards, research data repositories and current research systems (CRIS), and the open source tool DSpace is analyzed for the implementation of the research data repository prototype, this prototype is validated with its respective implementation.

## RESUMEN

Durante el desarrollo de procesos investigación científica se generan datos que describen el contexto de las investigaciones y también dan soporte a las investigaciones, estos datos pueden reutilizarse, siendo útiles para la generación de nuevas producciones intelectuales, por lo que es información relevante y se requiere implementar una solución software que ayude a automatizar la gestión de estos datos. Una solución software como lo es un repositorio de datos de investigación proporciona respuestas a las necesidades planteadas. Por esta razón, el presente trabajo tiene como objetivo el desarrollo de un prototipo de repositorio de datos de investigación para la publicación de conjuntos de datos vinculado a elementos de los actuales sistemas de información de investigación (CRIS - *Current Research Information System*) para la gestión de grupos de investigación de la Universidad del Cauca. Se realiza una fundamentación teórica sobre datos de investigación, estándares de metadatos, repositorios de datos de investigación y sistemas actuales de investigación (CRIS), y se analiza la herramienta de código abierto DSpace para la implementación del prototipo de repositorio de datos de investigación, este prototipo se valida con su respectiva implementación.

# 1. INTRODUCCIÓN

El auge de las Tecnologías de la información y las comunicaciones (TIC) en los últimos años ha llevado a la humanidad a migrar de un manejo de datos e información manual a la gestión y digitalización de esta información de manera electrónica. Este cambio ha generado múltiples beneficios entre los que se destacan la facilidad de acceso, almacenamiento eficiente de grandes cantidades de información, alto procesamiento y automatización de procesos entre otras. Este impacto trasciende en la incursión de tecnologías de información y comunicaciones en ámbitos como el de educación superior como un elemento fundamental, de hecho, la Organización de las Naciones Unidas para la Cultura, las Ciencias y la Educación (UNESCO) subraya la importancia de la integración de las TIC, partiendo de que son relevantes en el área de prácticas curriculares de innovación como en el acceso de la población a la educación superior (UNESCO, 1998). Lo anterior, brinda oportunidades para impulsar transformaciones en términos de la disponibilidad de los recursos para el aprendizaje, al igual que la orientación y apoyo de la gestión en lo que refiere principalmente a producción investigativa en beneficio de una universidad abierta. Entendida la universidad abierta, según Vinet y Zhedanov (2011) como “aquella que favorece el acceso abierto a la información, escucha a la comunidad universitaria, tomando decisiones según sus necesidades y prioridades, y genera espacios de colaboración para desarrollar servicios con todos los actores de la comunidad universitaria”.

Cabe resaltar que al hablar de universidad abierta y de datos abiertos, es importante tener en cuenta directrices locales sobre la gobernanza de estos datos como lo establecido en el Modelo de Gobernanza de la Infraestructura de Datos Colombiano (MGID) el cual presenta un “esquema de responsabilidad distribuida, donde intervienen diversos actores para promover la democratización de la infraestructura de datos con el fin de que los diversos actores del ecosistema de datos, incluida la ciudadanía, se vean beneficiados, y promueva el desarrollo en un contexto de creciente digitalización” (PolicyLab et al., 2022).

En el ámbito de la educación superior uno de los propósitos más importantes de las universidades es la investigación científica como pilar fundamental para el servicio de las mismas instituciones educativas y comunidad en general, siendo la investigación una estrategia que a juicio de Fullwood et al. (2013) permite generar conocimiento, destacando en ese mismo sentido también el aporte en crear conocimiento que se manifiesta en la divulgación y difusión mediante la publicación de ese conocimiento. Lo anterior origina un enriquecimiento en términos de adquisición y elaboración de nuevo conocimiento y a su vez según Kindling et al. (2017) “terceros pueden evaluar el conocimiento académico en función de estos datos” como activos de información en múltiples disciplinas, a fin de generar entre otras cosas, la posibilidad a los interesados de tomar una postura crítica frente a los conocimientos adquiridos o en su defecto construcción de un aprendizaje que

genere amplios beneficios a la sociedad, ya que como manifiestan Kindling et al. (2017), al ser “conjuntos de datos generados y seleccionados con precisión se pueden volver a analizar para validar los resultados de la investigación o se pueden reutilizar y readaptar para responder a diferentes preguntas de investigación”.

En ese sentido algunos de los mecanismos para promover este tipo de actividades correspondientes a investigación son de conformidad con Medina (2021) “observatorios, centros de pensamiento, centros de excelencia, escuelas de pensamiento entre otras estructuras organizativas”. De igual manera, de acuerdo con SECRETARIA - SENADO (2009), se puede resaltar los centros de productividad, los parques tecnológicos, los grupos de investigación.

En virtud de lo anterior, las universidades, como instituciones públicas, al tener necesidades de gestión de la información, en las cuales como lo afirma Fullwood et al. (2013) “la gestión del conocimiento en las universidades es muy limitada” y particularmente en información de procesos de investigación también se han visto enfrentadas a la necesidad de encontrar medios de gestión de información digital para poder centralizar esta información, y con ello compartirla para posteriores producciones de investigación que conlleven a un mayor impacto social.

En el contexto universitario, especialmente en los grupos de investigación se generan publicaciones científicas de relevancia, cabe enfatizar que estas producciones intelectuales se originan a partir de datos generados durante el proceso investigativo. Estos datos no solo describen el panorama de las investigaciones, es decir, el contexto de las investigaciones. Sin embargo, se evidencia que existen necesidades en términos de servicios e infraestructura para la preservación, apoyo, difusión y gestión de dichos datos generados en los procesos de investigación incluidos las producciones intelectuales como lo son las publicaciones con información necesaria que demuestre confiabilidad y pueda ser reutilizable.

Ante este escenario con fin de facilitar la reutilización de la información como herramienta de apoyo a la generación de futuras investigaciones y nuevos productos, surge la siguiente pregunta de investigación que pretende responderse con el desarrollo de este trabajo de grado:

***¿Cómo integrar un repositorio de datos de investigación con elementos de los CRIS para grupos de investigación en la Universidad del Cauca?***

## **1.1 Justificación**

A partir de las consideraciones de Vidotti et al. (2018) en un contexto académico “los repositorios de datos de investigación adquieren una gran relevancia, ya que, en la mayoría de los casos, permitirán el almacenamiento, la preservación y la difusión de los datos producidos”, y los repositorios de datos son una herramienta de gestión de información de investigación en la cual resalta la relevancia de los metadatos de información de investigación que “describen el contexto de los datos de investigación y las herramientas con las que se generaron, almacenaron, procesaron y analizaron” (Kindling & Schirmbacher, 2013), y a partir del manejo de estándares de estos metadatos ya mencionados, se logra la interoperabilidad para que la información se pueda compartir y que sea accesible de manera óptima.

Para esto existen múltiples soluciones software para la implementación de repositorios de datos para las necesidades planteadas anteriormente, tales como: DSpace, EPrints, Fedora y Greenstone, siendo estas soluciones de software un aporte significativo en el apoyo de la publicación y gestión de información de investigación ligado a CRIS en un entorno educativo universitario como es el de la Universidad del Cauca.

Evidentemente la Universidad del Cauca es una Universidad que maneja un alto volumen de producciones intelectuales, por consiguiente, necesidades de gestión tanto de la información de las producciones como de la información obtenida durante los procesos investigativos, necesidades que aún no están del todo resueltas ya que al explorar en el registro de repositorios de datos de investigación definido por (Re3data.org, n.d.) como “un registro global de repositorios de datos de investigación que cubre repositorios de datos de investigación de diferentes disciplinas académicas. Incluye repositorios que permiten el almacenamiento permanente y el acceso a conjuntos de datos para investigadores, organismos de financiación, editores e instituciones académicas.” La Universidad del Cauca no cuenta con un repositorio registrado.

## **1.2 Objetivos**

### **Objetivo general**

- Desarrollar un prototipo de repositorio de datos de investigación para la publicación de conjuntos de datos vinculado a elementos de los actuales sistemas de información de investigación (CRIS) para la gestión de grupos de investigación de la Universidad del Cauca.

### **Objetivos específicos**

- Caracterizar datos de investigación, estándares de metadatos, repositorios de datos de investigación y sistemas actuales de investigación (CRIS) aplicables a grupos de investigación.

- Definir un modelo para la publicación de datos de investigación que integre un repositorio de datos y los sistemas (CRIS) a partir de la caracterización realizada para un ámbito universitario.
- Evaluar el modelo de publicación de datos de investigación definido mediante la implementación de un prototipo de repositorio de datos utilizando DSpace-CRIS para un grupo de investigación de la Universidad del Cauca.

### **1.3 Aportes**

Con este proyecto se aporta significativamente al presentar una solución software que busca solventar las necesidades de gestión de la información de investigación para mantener y sumar esfuerzos para que la Universidad del Cauca pueda destacar en la producción investigativa no solo a nivel local, sino posicionarse internacionalmente, partiendo de los beneficios y apoyo que pueden significar la publicación y reutilización de información como ayuda a futuras producciones intelectuales que potenciarán provechos o resultados como la inversión y reconocimientos a todos los actores involucrados de la institución.

## 2. DE UN MAPEO SISTEMÁTICO A UNA REVISIÓN SISTEMÁTICA DE LA LITERATURA

En esta sección, se presenta un mapeo sistemático que siguió un rigor metodológico de investigación, el cual ofrece una perspectiva completa y estructurada de la literatura existente. Este enfoque sistemático permitió la identificación y también analizar las investigaciones previas, proporcionando una base sólida para comprender el estado actual del conocimiento en el área de estudio. El objetivo fue buscar, organizar y clasificar la información relevante, resaltando tendencias, brechas y enfoques fundamentales presentes en la literatura científica. Al final de esta sección se presentan resultados de síntesis propios de una revisión sistemática que van más allá de los resultados de un mapeo sistemático.

### 2.1 Procedimiento metodológico

De manera detallada y sistemática se presenta la estrategia junto con las etapas seguidas para llevar a cabo este mapeo sistemático. A continuación, se describen las herramientas, técnicas y métodos utilizados en la concepción y ejecución del estudio, proporcionando una guía sobre cómo se recopilaron, analizaron e interpretaron los datos.

#### 2.1.1 Cadena de búsqueda

La fuente de investigación seleccionada para realizar la búsqueda de publicaciones y revisión de la literatura utilizada en este proyecto fue la base de datos científica SCOPUS, la elección de esta base de datos se sustenta en que “proporciona una visión general completa de la producción de investigación en todo el mundo en los campos de la ciencia, la tecnología, la medicina, las ciencias sociales y las artes y las humanidades” (Elsevier, n.d.-a). Además de contener investigaciones relevantes y acreditadas, proporcionando acceso a datos confiables (Elsevier, n.d.-b). Asimismo, Scopus tiene 26.591 revistas activas revisadas por pares y 192 revistas especializadas (Elsevier, 2020). La finalidad será comprender el panorama de las temáticas de esta investigación y convertir los resultados en insumo para definir las características a considerar para el desarrollo de los objetivos.

La cadena de búsqueda se elabora a partir de cinco términos, los cuales se consideran relevantes para las necesidades de este proyecto, los términos "*research data*" y "*research data repository*" se agrupan en un "AND", y los términos "*Current Research Information System*", "CRIS" y "*metadata standard*" con el operador lógico "OR", posteriormente a los resultados obtenidos a partir de la cadena de búsqueda, se aplican criterios de inclusión y exclusión con el fin de filtrar los documentos obtenidos inicialmente.

La cadena de búsqueda inicial sin filtros arrojó un total de 4.091 documentos, a estos resultados se le aplican los siguientes criterios de inclusión y exclusión.

### **2.1.2 Criterios de inclusión y exclusión**

Los criterios de inclusión y exclusión presentan una serie de condiciones, estas condiciones son las que filtran aquellos documentos relevantes obtenidos con la cadena de búsqueda inicial, luego de aplicar este filtro se obtiene como resultado un número de documentos reducido que cumple con las características que se consideran relevantes para la temática de la investigación, estos criterios son definidos a continuación.

#### **Criterios de inclusión:**

Únicamente se considerarán los documentos publicados desde el año 2017 hasta la actualidad (2023), además aquellos que estén escritos en inglés o español, igualmente se considerarán las publicaciones disponibles a través de la suscripción de la institución Universidad del Cauca, también se incluyen solo los documentos de tipo artículo y que sean de acceso abierto. Esto se realiza con el objetivo de obtener un panorama reciente de las temáticas a buscar.

#### **Criterios de exclusión:**

Se descartan los documentos que estén escritos en un idioma diferente a inglés o español, la elección del idioma inglés se basa en la disponibilidad de un mayor número de documentos y en español por ser un idioma en el cual se puede encontrar documentos en un contexto latino, también se excluyen los artículos que no contengan al menos uno de los siguientes términos "*Research Data*", "*Research Data Management*", "*Metadata Standards*".

### **2.1.3 Criterios para la evaluación de la calidad**

Una vez aplicados los criterios de inclusión y exclusión se obtuvieron un total de 170 documentos a los cuales se les aplica los criterios de calidad, lo cual consistió en realizar la revisión de la introducción de los artículos para valorar la relevancia de la temática, alineación con los objetivos de la investigación, actualidad y pertinencia. Dado el caso de que el documento no se ajustara a las necesidades de la investigación o el contexto expuesto en la introducción de la publicación se enfoca en temáticas irrelevantes para la investigación se excluye de la selección final.

En la Tabla 1 se presentan los artículos finales luego de aplicar los criterios previamente descritos.

*Tabla 1. Documentos obtenidos al aplicar criterios para la evaluación de la calidad a la cadena de búsqueda en la base de datos especializada Scopus.*

#	Referencia	#	Referencia
1	(Bhardwaj, 2019)	56	(Islam et al., 2020)
2	(Ishak et al., 2019)	57	(Poblet et al., 2018)
3	(Kindling & Strecker, 2022)	58	(Pronk, 2019)
4	(Hansson & Dahlgren, 2022)	59	(Cousijn et al., 2019)
5	(Nie et al., 2021)	60	(Quarati & Raffaghelli, 2022)
6	(Downs, 2021)	61	(de la Hidalga et al., 2022)
7	(Pereira et al., 2019)	62	(Marlina et al., 2022)
8	(Arend et al., 2020)	63	(Bloemers & Montesanti, 2020)
9	(Latif et al., 2019)	64	(Cho & Yu, 2018)
10	(Rumble et al., 2019)	65	(Wasner et al., 2018)
11	(Wiljes & Cimiano, 2019)	66	(Calhoun et al., 2019)
12	(Nishida et al., 2020)	67	(Graf et al., 2020)
13	(Lefebvre et al., 2020)	68	(Iglezakis, 2020)
14	(Schirwagen et al., 2019).	69	(Stocker et al., 2020)
15	(Oliver et al., 2020)	70	(Suhr et al., 2020)
16	(Hrynaskiewicz et al., 2020)	71	(Ekström, 2023)
17	(Cerda-Cosme & Méndez, 2022)	72	(Jiao & Li, 2022)
18	(Lefebvre, 2023)	73	(Van De Sandt et al., 2019)
19	(Bossaller & Million, 2022)	74	(Klump et al., 2021)
20	(Buchholz, 2021)	75	(Borycz, 2021)
21	(Korkeamäki et al., 2022)	76	(Austin et al., 2021)
22	(Latif et al., 2021)	77	(Curdt et al., 2022)
23	(Yan et al., 2022)	78	(Adriaan et al., 2020)
24	(Bellgard, 2020)	79	(Cruz et al., 2019)
25	(Labastida & Margoni, 2020)	80	(Mahony, 2022)
26	(Posavec et al., 2020)	81	(Jackson, 2021)
27	(Minamiyama et al., 2020)	82	(Thessen et al., 2019)
28	(Defranco & Voas, 2021)	83	(Miksa et al., 2021)
29	(Zwölf et al., 2019)	84	(Hettne et al., 2020)
30	(Masinde et al., 2021)	85	(Zuiderwijk & Spiers, 2019)
31	(Fitschen et al., 2019)	86	(Schöpfel et al., 2018)
32	(Manik et al., 2022)	87	(Schopf et al., 2022)
33	(Hedeland, 2020)	88	(K. M. Gregory et al., 2020)
34	(Töwe & Barillari, 2020).	89	(Jejkal et al., 2022)
35	(Mannheimer et al., 2019)	90	(Bezuidenhout et al., 2021)
36	(Schröder et al., 2022)	91	(Nitecki & Alter, 2021)



37	(Rousi & Laakso, 2020)	92	(Aleixandre-benavent et al., 2020)
38	(Hashim, 2019)	93	(Cox et al., 2018)
39	(Lehmann et al., 2023)	94	(Subirats-coll et al., 2022)
40	(Enescu et al., 2018)	95	(Schmieg et al., 2022)
41	(Grasse et al., 2018)	96	(Gonzales et al., 2022)
42	(K. Gregory et al., 2018).	97	(McKeague et al., 2020)
43	(Mushi et al., 2020)	98	(Bahim et al., 2020)
44	(Cox et al., 2019)	99	(David et al., 2020)
45	(Koltay, 2020)	100	(Tang & Hu, 2019)
46	(Brandt et al., 2022)	101	(Farinelli & Zigoni, 2022)
47	(Hasselbring et al., 2020)	102	(Jetten et al., 2019)
48	(Štebe et al., 2020)	103	(Rosalina Vázquez, 2022)
49	(Grunzke et al., 2019)	104	(Zervas et al., 2019)
50	(Shipman & Tang, 2019)	105	(Biesenbender et al., 2019)
51	(Senagi & Tonnang, 2022)	106	(Schöpfel et al., 2017)
52	(Brandt et al., 2021)	107	(Rousi, 2022)
53	(Xu et al., 2022)	108	(Sivertsen, 2019)
54	(Kern & Hienert, 2018)	109	(Dvořák et al., 2019)
55	(Medina-Smith et al., 2021)	110	(Fabre & Egret, 2020)

Como se puede observar en la Tabla 1, de los resultados obtenidos totales y una vez aplicados los criterios de evaluación de la calidad, se obtienen un total de 110 resultados, los cuales se encuentran organizados mediante un identificador por artículo y su respectiva referencia.

## 2.2 Resultados

De los documentos seleccionados se obtienen estadísticas que pueden servir de recurso para el consolidado del mapeo sistemático, estas estadísticas son presentadas a continuación.

### 2.2.1 Artículos por año

La Figura 1 muestra de forma gráfica la distribución de la cantidad de artículos por año de los artículos seleccionados para la revisión literaria en el rango definido del periodo entre los años 2017 y 2023. Cabe destacar que entre los años 2018 y 2022 se produjo un notable incremento de artículos y el año donde mayor contribución de publicaciones se observa es el año 2020.

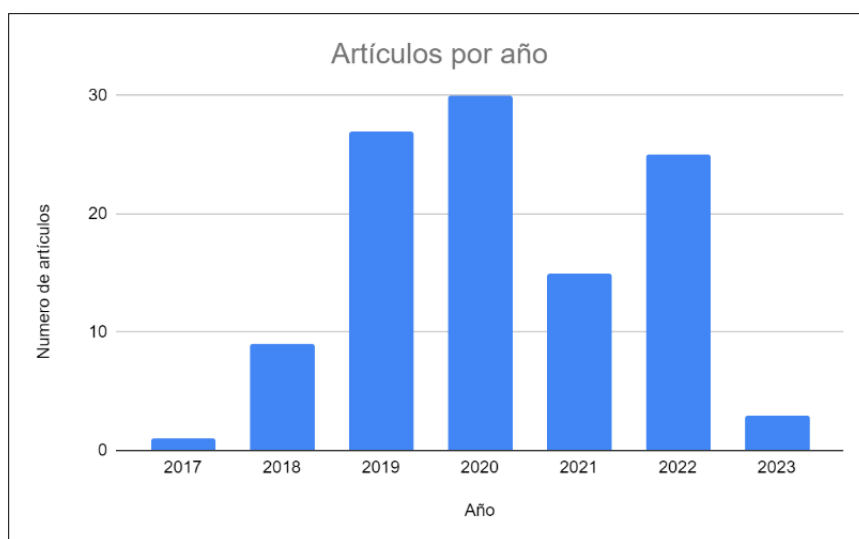


Figura 1. Número de artículos por año.

### 2.2.2 Licencias y derechos de autor

En el mapeo sistemático, de los 110 artículos recopilados, se detallan las distintas licencias de derechos de autor utilizadas. La Tabla 2 proporciona información específica sobre cada licencia, incluyendo el nombre, el número de artículos asociados y sus características distintivas. Este resumen facilita una visión rápida y detallada de la información esencial relacionada con las licencias encontradas.

Tabla 2. Nombres de licencias de derechos de autor.

Licencia	Número Artículos	Característica
<i>Creative Commons</i>	15	Grupo de licencias, gratuitas y flexibles para especificar términos de uso de obras según necesidades.
Dominio Publico	5	Mas que una licencia trata de derechos de autor que han expirado, no son reclamados o se ha renunciado a ellos.
<i>Apache Software License</i>	2	Gratuita y usada para software
<i>OGLOpen Government License</i>	2	Gratuita y usada con fines de gobierno
GNU	2	Gratuita y usada en software
Artículos que mencionan relevancia de uso claro de licencias	21	

Como se observa en la Tabla 2 se encuentra que 21 de ellos mencionan o hacen referencia a algún tipo de licencia con relación a datos de investigación, representando un 23.1%, de los 21 artículos, 15 describen el uso de *Licencias Creative Commons* entre las cuales se destacan CC0 (*Creative Commons Zero*), CC BY (*Creative Commons Attribution*). En segundo lugar, se encuentran cinco (5) artículos que mencionan la licencia de dominio público, seguida de Apache Software License con dos (2) artículos, OGL con dos (2) artículos y GNU con dos (2) artículos.

La Figura 2, basada en la información detallada en la Tabla 2, ofrece una representación visual de las licencias y derechos de autor mencionados en los artículos revisados.

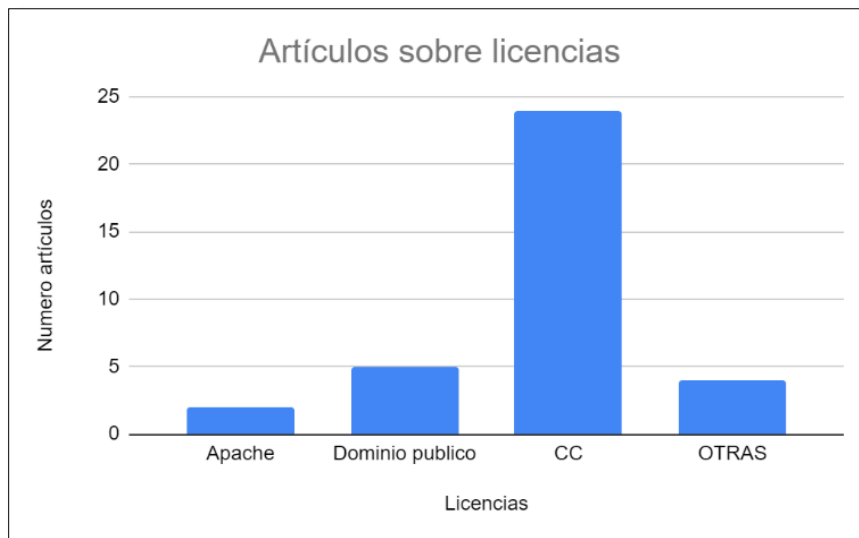


Figura 2. Cantidad de artículos por cada licencia de derechos de autor.

Es evidente que en la Figura 2, las licencias *Creative Commons*, especialmente CC0 (*Creative Commons Zero*) y CC BY (*Creative Commons Attribution*), dominan la preferencia entre los investigadores, representando la mayoría de las menciones con un total de 24 artículos.

### 2.2.3 Estándares de metadatos, normas y modelos

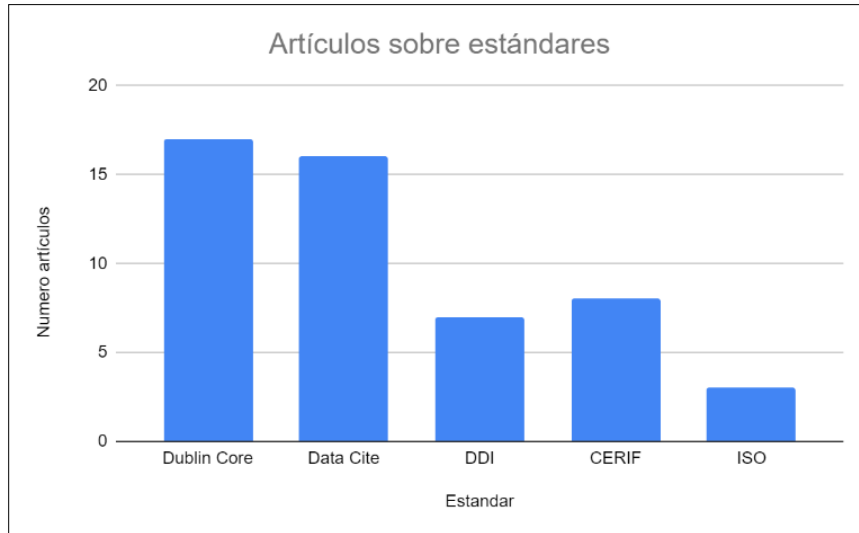
En la Tabla 3 se detallan los estándares de metadatos, destacando el nombre específico de cada estándar, su frecuencia de aparición en los artículos revisados y las características distintivas asociadas. Esta presentación brinda una visión rápida y completa de la presencia y particularidades de los estándares de metadatos abordados en la revisión sistemática.

Tabla 3. Nombre de estándares de metadatos y total de publicaciones.

Estándar de metadatos	Número de artículos que lo mencionan	Característica	Gratuito
Dublín Core	17	Describir recursos digitales	Si
DataCite	16	Publicación y citación de datos de investigación e identificadores persistentes	Si
CERIF	8	Soporte CRIS	Si
DDI	7	Describir datos cuantitativos y cualitativos	Si
Normas ISO	3	Descripción de recursos digitales	Algunos
Artículos que mencionan uso de estándares de metadatos	36		

De los 110 artículos revisados, un total de 36 hacen mención o referencia a estándares de metadatos, normas o modelos, representando el 39.6% del conjunto. Entre estos artículos, los estándares más frecuentemente citados son Dublín Core, mencionado en 17 artículos, seguido de Datacite, también con 17 artículos, y el estándar DDI, presente en 7. Asimismo, el modelo CERIF es referenciado en 8 artículos, y se hacen alusiones a normas ISO en 3 artículos.

La Figura 3, a continuación, resume la presencia y distribución de estándares de metadatos, normas y modelos en los 110 artículos revisados. Este análisis ofrece una visión clara de la frecuencia con la que se hacen referencias a diferentes estándares en la literatura.



*Figura 3. Cantidad de publicaciones que abordan estándares de metadatos.*

En la Figura 3, Dublín Core y Datacite emergen como los estándares más citados, ambos mencionados en 17 y 16 artículos respectivamente, subrayando su prominente posición en la comunidad académica. La presencia de estándares como DDI, CERIF y las normas ISO, aunque en menor frecuencia, indica una diversidad en la adopción de enfoques para la gestión y descripción de datos de investigación.

#### **2.2.4 Repositorios de datos de investigación**

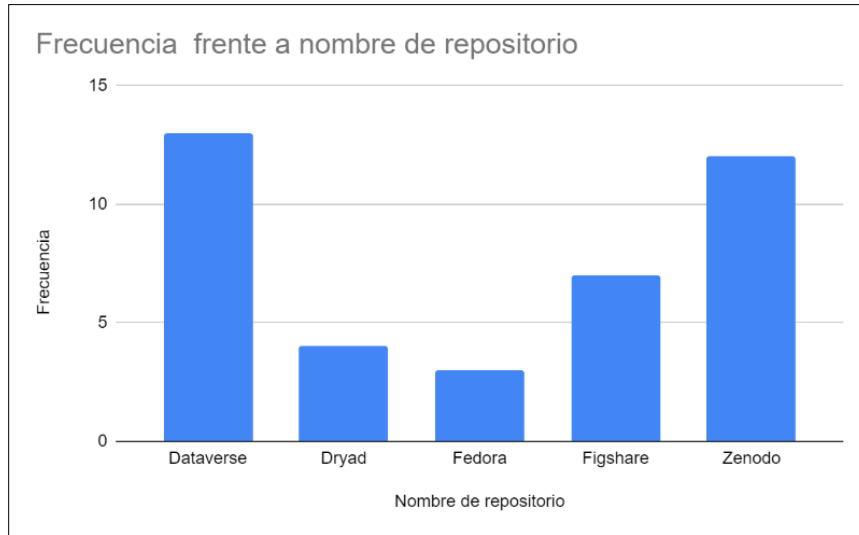
La Tabla 4 destaca repositorios relevantes y su frecuencia de uso en los artículos seleccionados, proporcionando una visión clara de su prevalencia. Además, en dicha Tabla, la presencia de "limitado" en la columna "gratuito" indica restricciones de acceso y posibles costos adicionales para alojamiento, infraestructura y mantenimiento, según requisitos específicos. A continuación, se presentan los repositorios encontrados en el 42% de las publicaciones en las que se manifiesta explícitamente el nombre del repositorio.

Tabla 4. Repositorios mayormente descritos en revisión bibliográfica.

Repositorios	Número de artículos	Multidisciplinario	Gratuito	Especializa	Código abierto
Dataverse	13	SI	SI	Ciencias sociales, salud.	SI
Dryad	4	SI	Limitado	Biología, ecología, genética, medicina.	SI
Fedora	3	SI	Limitado		SI
Figshare	7	SI	Limitado	Ciencias naturales, ciencias sociales, humanidades, ciencias de la ingeniería.	NO
Zenodo	12	SI	SI	Incluye ciencias naturales, ciencias sociales, humanidades, ingeniería, medicina, tecnología, entre otros.	SI
Artículos que mencionan o abordan sobre repositorios de datos de investigación.	33				

En la Tabla 4, se observa que, de las 110 publicaciones revisadas, el término repositorio de datos aparece en aproximadamente el 72% de ellas, mientras que en el 28% restante no se aborda este tema. Dentro de las 79 publicaciones relacionadas con repositorios de datos de investigación, en el 58% de las publicaciones no menciona explícitamente un repositorio específico, mientras que en el 42% de estas publicaciones se especifica un nombre de repositorio de investigación en particular.

En la Figura 1, se proporciona una representación visual sobre el nombre de los diferentes repositorios de datos de investigación en los artículos revisados. Este análisis permite apreciar la proporción de publicaciones que abordan los repositorios de datos.



*Figura 4. Cantidad de publicaciones que abordan repositorios de datos de investigación.*

En la Figura 4, además, se destaca que los repositorios que registraron la mayor frecuencia de menciones en los artículos revisados son Dataverse, Zenodo y Figshare. Estos repositorios emergen como los más recurrentes en la literatura examinada, sugiriendo su significativa presencia y relevancia en el contexto de la gestión de datos de investigación.

Además, durante la revisión bibliográfica se identificaron plataformas que facilitan la creación de repositorios de datos de investigación, dichas plataformas se muestran en la Tabla 5 y en la Figura 5.

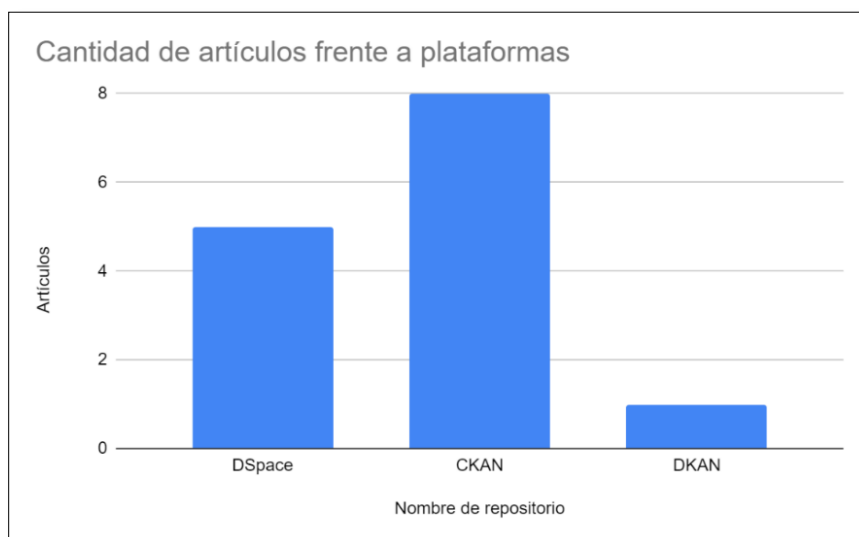
*Tabla 5. Plataformas para crear repositorios de datos de investigación*

Plataforma Software	Número de artículos	Es repositorio	Gratuito	Código abierto	Área
Invenio	3	NO	Limitado	SI	Gestión y distribución de contenido digital, como publicaciones académicas, conjuntos de datos, archivos multimedia, tesis, informes de investigación, entre otros.
DSpace	5	SI	SI	SI	Gestionar y preservar materiales digitales, como artículos, tesis, informes, imágenes y otros tipos de contenido académico y de investigación.
CKAN	8	NO	Limitado	SI	Gobierno abierto, ciencia abierta, datos geospaciales, datos de investigación, datos

					medioambientales, datos de salud, entre otros.
--	--	--	--	--	--

En la Tabla 5 se destacan las plataformas utilizadas para crear repositorios, indicando el nombre de cada plataforma y la frecuencia con la que se abordaron en los artículos seleccionados. Esta síntesis ofrece una visión directa y detallada de las plataformas relevantes en la revisión sistemática.

A continuación, se presenta en la Figura 5, el detalle de las plataformas más utilizadas en el contexto de este mapeo sistemático.



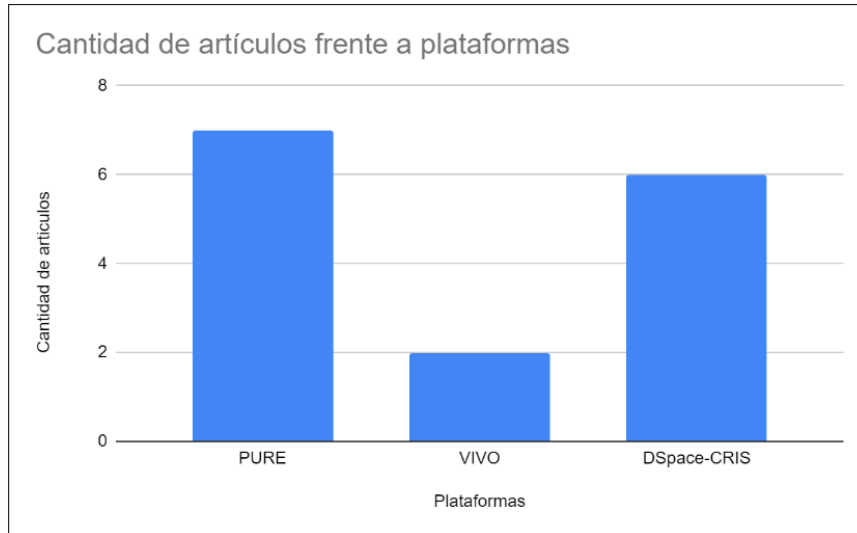
*Figura 5. Cantidad de publicaciones que abordan plataforma para creación de repositorios de datos de investigación.*

La Figura 5 muestra que Ckan y DSpace son las plataformas más destacadas en la literatura examinada, resaltando su importancia y popularidad en creación y la gestión de repositorios de datos. Esto sugiere que estas plataformas son ampliamente preferidas y utilizadas en la comunidad académica para abordar las necesidades de manejo y compartición de datos de investigación.

### **2.2.5 Current Research Information System (CRIS)**

De las 110 publicaciones revisadas, se identificó que el término "*Current Research Information System*" (CRIS) aparece en 10 de ellas. De este conjunto, 8 publicaciones abordan explícitamente sistemas o plataformas relacionadas, como se detalla en la Figura 6





*Figura 6. Cantidad de publicaciones por cada plataforma CRIS.*

La Figura 6 presenta la distribución de publicaciones según la plataforma CRIS en la revisión bibliográfica. Con esto, se evidencia que las plataformas más frecuentemente mencionadas son PURE y DSpace-CRIS. Estas dos plataformas sobresalen, indicando su popularidad en el contexto de los Sistemas de Información de Investigación y Gestión de Investigación (CRIS). Lo anterior sugiere que PURE y DSpace-CRIS son opciones significativas y bien reconocidas en el ámbito de la gestión de información y actividades de investigación.

### **2.3 Hacia una revisión sistemática**

A partir de los estudios seleccionados y los resultados y análisis presentados, posteriormente se revisaron exhaustivamente en la búsqueda de aportes relevantes para esta investigación, por lo cual, se organizan los diferentes elementos distintivos de los artículos por temáticas, los cuales se describen en la siguiente sección, con la finalidad de establecer un consolidado que defina a partir de los resultados de la revisión, una síntesis de los aspectos principales a considerar para la implementación del prototipo propuesto en este trabajo.

#### **2.3.1 Datos de investigación**

Los datos de investigación de acuerdo con la OECD (2007) son aquellos “utilizados como fuentes primarias para la investigación científica, y que son comúnmente aceptados en la comunidad científica como necesarios para validar los resultados de la investigación”. Esta definición también se respalda en la definición proporcionada por la Comisión (2020), la cual establece que un dato de investigación “se refiere a la información, en particular hechos o cifras, recopilada para ser examinada y considerada como base para el razonamiento, el debate o el cálculo” (Commission, 2020). En concordancia con esto, Ishak et al. (2019) En su

estudio encuentra que los datos de investigación son cruciales; aparte de ser datos recolectados para analizarlos, son el reflejo o resultado de un estudio, estos datos pasan a transformarse en nuevo conocimiento y, después, ser un componente esencial para promover el intercambio o apoyar a otros investigadores. De forma similar, el estudio llevado a cabo por Nishida et al. (2020) evidenció que los datos de investigación son un respaldo que permite validar los resultados de una investigación y, en ese sentido, son activos aceptados por una comunidad científica, por ende, resulta fundamental que se registren de una manera completa y con la mayor precisión posible.

Se debe subrayar que, a lo largo de un proceso de investigación, se generan o recolectan cada vez un mayor volumen y diferentes tipos de datos o información, esta información corresponde a los datos de investigación que tienen la característica principal de ser soporte del proceso investigativo al cual están relacionados, como se menciona en Yan et al. (2022). Se trata, por tanto, de varios tipos de datos como pueden ser experimentales, encuestas, entrevistas, casos de estudio, videos, audios, muestras, datos de observación personal, datos web, estadísticos y datos de simulación, entre otros, los cuales se obtienen mediante recopilación, observación o análisis y fueron recolectados en un proceso de investigación como se expone en Manik et al. (2022). Toda esta variedad de datos puede estar en su estado inicial o con estructura, es decir, ser información procesada, limpia y organizada, puede incluir recursos como textos, algoritmos, gráficos, conforme a lo indicado por Hrynaskiewicz (2020) y Hashim (2019).

Según lo expuesto, se puede entender que un dato de investigación abarca cualquier tipo de dato producido o recolectado digitalmente, o que en su defecto se generó para convertirse en información analizada o utilizada en un contexto de un proceso o proyecto de investigación científica. Un dato de investigación puede comprender desde datos primarios, es decir, datos que fueron generados durante la investigación científica o también pueden ser datos secundarios lo que significa que son datos que fueron recolectados por otros investigadores y que se emplearon para algún contexto de investigación, según lo señalado por Hrynaskiewicz et al. (2020).

El objetivo de los datos de investigación es servir como un recurso esencial para otros investigadores interesados en el campo específico al cual pertenece dicho dato de investigación, con lo cual se aumente y promueva la colaboración científica, la verificación de resultados obtenidos, se respalde la toma de decisiones en nuevas investigaciones, apoyándose en estos datos para la generación de nuevo conocimiento y se acelere el proceso de investigación, así también lo afirma Štebe et al. (2020), al señalar que estos datos después tendrán un intercambio amplio y así se beneficiarán financiadores, comunidades científicas de diferentes disciplinas y promoverán la innovación y la investigación de la comunidad, no obstante, reconoce que en el intercambio de datos existen dificultades como la falta de

conocimiento en la gestión de datos, datos efímeros, temor a presentar datos de baja calidad. El mismo desafío o dificultades en el intercambio de estos datos digitalizados, se describe en Islam et al. (2020).

Otro desafío relacionado con los datos de investigación es la interdisciplinariedad, que se puede ver en Cho y Yu, (2018), en donde se afirma que, en lo referente a los datos de las publicaciones científicas, una gran mayoría de artículos se apoyan o apuntan referenciando disciplinas o áreas de estudio diferentes a la disciplina principal del artículo en cuestión, es decir, que por parte de los investigadores o autores se tiende a recurrir por apoyarse en estudios de diversos campos para respaldar sus investigaciones. Asimismo, en sus artículos, K. M. Gregory et al. (2020) y Mahony (2022) resaltan los datos abiertos de investigación como activos que tienen un gran potencial para mejorar la productividad, eficiencia y reproducibilidad de la ciencia, aunque se admite la existencia de desafíos debido a la diversidad en disciplinas y usuarios. Por ejemplo, indican los autores que un investigador puede necesitar o producir datos para diversas finalidades. Sin embargo, hay un fuerte interés en que se promueva el intercambio y la reutilización de datos, por ello, tanto los agentes financiadores como las organizaciones generan políticas para superar estas dificultades. También destacan que es importante subrayar que los investigadores generan datos, sino que pueden requerir de datos producidos y publicados por otros investigadores y que en ese caso los datos pasarían por procesos de utilización y respectivo análisis para darle más sentido a los datos. Además, los autores supracitados señalan que la ausencia de estandarización en la manera de cómo se hace la recopilación, descripción e intercambio de datos puede dificultar las tareas de búsqueda, evaluación y reutilización de estos.

En relación con interdisciplinariedad, en el estudio de Jiao y Li (2022), se resalta que hay un crecimiento en los repositorios encaminados a ser de índole interdisciplinar, también sostienen que los datos son uno de los recursos esenciales cuando se trata de investigación científica y enfatizan que los patrocinadores actualmente exigen el intercambio de datos para promover la disponibilidad de los mismos.

### **2.3.1.1 Principios FAIR**

En el ámbito de calidad de datos en Kindling y Strecker (2022) se hacen referencia a que “En el contexto de los datos, la calidad a menudo se conceptualiza como la idoneidad para el uso”, esto va de la mano con los principios denominados FAIR, acrónimo de (*Findable, Accessible, Interoperable y Reusable*). Dichos principios FAIR son de conformidad con Wilkinson (2016) una serie de principios puntuales y cuantificables cuyo núcleo se centra en “en mejorar la capacidad de las máquinas para encontrar y usar automáticamente los datos, además de apoyar su reutilización por parte de las personas”. Austin et al. (2021) enfatizan que los principios FAIR buscan mejorar la infraestructura que permite acceder y reutilizar los datos de

investigación para que sean más accesibles y útiles, pretenden ser una referencia para aquellos que quieran compartir sus datos con otros para su reutilización.

Estos principios son ampliamente referenciados en los artículos estudiados, como en Minamiyama et al. (2020), donde se plantea que pueden existir costos adicionales para interpretar o reutilizar datos al no definir claramente condiciones de uso de datos y no utilizar normas comunitarias para compartir y publicar datos como lo son los principios FAIR, también en Cerda-Cosme y Méndez (2022) donde se alude que los principios FAIR se centran en los componentes técnicos que harían que un conjunto de datos sea localizable, accesible, interoperable o reutilizable, los mismos autores plantean nueve atributos como alternativa para medir FAIR, los cuales son disponibilidad, accesibilidad, formato, licencia, enlace, financiación, política editorial, contenido y estadísticas, cabe resaltar que en Brandt et al. (2022) se refiere también sobre la importancia de los principios FAIR y muestra la relevancia del uso de estándares de metadatos para lograr cumplir con estos principios. A menudo la importancia de los principios FAIR se destaca por ser un factor significativo en la legibilidad e interoperabilidad de las máquinas, no obstante, como se ha venido mencionando estos principios están destinados a servir de pautas para facilitar la legibilidad por parte de los humanos y permitir la posterior reutilización de la información; existe también otros principios que mejoran aspectos como la accesibilidad e interoperabilidad como lo son los principios de datos abiertos enlazados (Ganske et al., 2020). Por lo tanto, el intercambio de datos de investigación siguiendo los principios FAIR generan valor agregado en los avances de la ciencia lo que a su vez contribuye a una mayor calidad en las producciones científicas.

En un proceso de gestión de datos de investigación entonces, adquiere importancia el conjunto de principios que se han creado para fines de promover la gestión efectiva de los datos de investigación, los ya mencionados principios FAIR, sobre los cuales Wilkinson (2016) afirma que “no sugieren ninguna tecnología, estándar o solución de implementación específica; además, los principios no son, en sí mismos, un estándar o una especificación”.

De igual manera, Bahim et al. (2020) percibe que los principios FAIR no especifican cómo lograr que los datos cumplan con estos principios, pero se han desarrollado diferentes métodos y herramientas para implementarlos. Algunas de estas herramientas se centran en evaluar los datos, mientras que otras se centran en evaluar el plan de gestión de datos. Sin embargo, se ha prestado menos atención al tema de la accesibilidad y se ha otorgado importancia al principio de encontrar los datos. Por ende, los autores proponen un modelo de madurez de datos FAIR que identifica todos los elementos de FAIR que se pueden analizar. Este modelo se publica como una recomendación de *Research Data Alliance* (RDA) y se sugiere que se utilice en la elaboración de planes de gestión de datos de investigación.

Se denota un consenso con los principios FAIR, sin importar la disciplina, son una herramienta fundamental para mejorar la gestión de datos científicos, pues sirven de guía para garantizar que los datos científicos sean accesibles, comprensibles y utilizables en el contexto investigativo. Esto puede contribuir a acelerar el progreso científico y promover una mayor colaboración entre investigadores, como se evidencia en los artículos mencionados anteriormente, así como en los trabajos de Zwölf et al. (2019), Fitschen et al. (2019), Rousi & Laakso (2020), Lehmann et al. (2023), Hasselbring et al. (2020), Downs (2021), Arend et al. (2020), Senagi y Tonnang (2022), Enescu et al. (2018), Brandt et al. (2021), Quarati & Raffaghelli (2022), Marlina et al. (2022), Iglezakis (2020), Suhret al. (2020), Austin et al. (2021), y Adriaan et al. (2020).

### **2.3.1.2 Gestión de datos de investigación (RDM)**

La gestión de datos de investigación (GDI) o *Research Data Management* (RDM) que según Lehmann et al. (2023) “tiene como objetivo principal capturar datos para allanar el camino hacia nuevos conocimientos científicos a largo plazo”. Suhr et al. (2020) también respaldan la idea de que RDM está dirigido a recopilar datos de investigación y soportar el flujo de todos los pasos que van desde la planificación hasta lo que implica el intercambio y respectiva publicación de los datos de investigación. RDM involucra un conjunto de prácticas que garantizan que los datos de investigación sean entre otras cosas, confiables, precisos, accesibles e impedir posibles pérdidas de la valiosa información como lo son los datos de investigación, con RDM se busca que exista un almacenamiento y conservación de los datos recolectados en un proyecto de investigación científica, por lo tanto, RDM es una pieza clave en lo concerniente al ciclo de vida de los datos según Marlina et al. (2022). Con RDM se busca gestionar y ser eficiente en preservar los datos en aras de potencializar lo que viene siendo la reutilización y reproducibilidad de estos activos por parte de comunidades científicas, esto se puede evidenciar en diferentes artículos revisados, como por ejemplo, en Lefebvre (2023) y en Wiljes y Cimiano (2019) donde se menciona que “Una gestión responsable de los datos de investigación es un requisito esencial de la buena práctica científica y proporciona una base sólida para una investigación excelente.”, a partir de esto se puede apreciar la trascendencia de la gestión de datos de investigación, que como anteriormente se indicó abarca el ciclo de vida de los datos, tal cual se menciona en Bhardwaj (2019).

En ese contexto, Bellgard (2020) también agrega que entre algunos procesos de investigación se pueden generar muchas fases, en donde cada fase origina sus propios datos, para los cuales se debe considerar el ciclo de vida en las diferentes fases que se originen y bajo este panorama como consecuencia, se requiere de diferentes enfoques de gestión de datos de investigación. GDI aborda la curaduría, recopilación, descripción, preservación, intercambio, difusión, reutilización y retención de datos que fueron generados durante todo el transcurso de un proceso

investigativo. Por su parte, Buchholz (2021) también afirma la trascendencia de la gestión de datos de investigación cuando expone que este aspecto se ha convertido en algo a resaltar dentro de lo que considera una buena práctica científica. En consecuencia, debe ser o establecerse como una práctica o parte integral en todo proceso investigativo. Es claro, por tanto, que una buena gestión de datos de investigación trae consigo o conlleva a poder innovar en el conocimiento, así como también poder luego integrar tal conocimiento y en muchos casos ser un factor trascendente en lo que refiere a la reutilización evitando, por ejemplo, el proceso lento de recopilación de datos primarios, evidentemente todo esto luego de que los datos hayan culminado su proceso de publicación según Wilkinson (2016) y Kern y Hienert (2018). Adicionalmente, las organizaciones de financiación de la investigación (RFO) exigen cada vez más que los autores de investigaciones entreguen y publiquen los datos de soporte o generados de sus investigaciones con la menor cantidad de restricciones posible, siempre respetando las responsabilidades éticas y legales, para contribuir a investigaciones futuras y que el retorno de las inversiones sea eficiente promoviendo la reutilización en lugar de crear nuevos conjuntos de datos, esto según lo expuesto en Bloemers y Montesanti (2020) y en Van De Sandt et al. (2019).

En Miksa et al. (2021) se aborda un tema estrechamente vinculado a los datos de investigación, que son los planes de gestión de datos (DMP). Estos planes son documentos que por lo general van acompañados de la propuesta o resultados de una investigación y su propósito es describir diversos aspectos relacionados con los datos utilizados, como su origen, almacenamiento, licencias, entre otros aspectos, por lo que resulta una descripción clave que brinda un respaldo fundamental a los datos de investigación y proporciona información de la gestión de datos durante el ciclo de vida de los datos. Asimismo, cabe destacar que en el trabajo de Gonzales et al. (2022) se presentan reglas sencillas que deberían considerarse al elaborar un plan de gestión e intercambio de datos de investigación, para optimizar la preservación y el intercambio de dichos datos.

En algunos de los trabajos analizados se menciona la importancia de un administrador de datos, a quien definen como una persona que es la responsable de recopilar y gestionar los datos, es decir que evalúe, por ejemplo, el cumplimiento de los estándares de metadatos, por ende, el administrador de datos actúa como un intermediario entre los investigadores y todo lo referente a los servicios de apoyo en los procesos de investigación (Posavec et al, 2020). El proceso de gestión de datos de investigación (RMD) también es un factor que se encuentra presente en algunos artículos que se revisaron dentro de los cuales se aborda que este tipo de prácticas se requieren en favor de una integridad de la investigación y por las preocupaciones por parte de los financiadores en su interés por que haya un crecimiento en la disponibilidad de datos abiertos y facilitar para que sean accesibles y sirvan de ejes centrales en fases como la reutilización de los datos, los cuales fueron producidos con financiación de fondos públicos y lo que se pretende

es que mediante la reutilización se logre generar e incentivar nuevo conocimiento (Enescu et al., 2018) (Cox et al., 2019).

En el proceso de gestión de datos también se destaca la importancia de formar o capacitar a los bibliotecarios con el fin de que adquieran conocimientos en temas como “curaduría de datos, requisitos disciplinarios/financiadores, planes de gestión de datos, almacenamiento de datos, infraestructura, seguridad, software, ética, consultas/instrucción/divulgación y metadatos.” (Shipman & Tang, 2019) y en sí, buenas prácticas con respecto a RMD. Marlina et al. (2022) también concuerda sobre la preparación o capacitación de los bibliotecarios en pro de mejorar las competencias con relación a RMD. Además, Austin et al. (2021) apoya la idea de que cualquier iniciativa que involucre la gestión de datos tiene como consecuencia, un potencial para que los datos de investigación tengan un valor agregado, esto ya sea en el proceso de ejecución del proyecto de investigación científica como una vez finalizado, incluso recalca el poder transmitir a las comunidades científicas los beneficios que surgen a partir de RMD como es el caso de una mejor reutilización y un índice a la alza en relación a aumento de citas.

Entre algunos artículos revisados se encuentra factores en común, los cuales consisten en la reutilización, así como la reproducibilidad, la transparencia, la conservación de datos de investigación y con ello también proporcionar disponibilidad o acceso a dichos recursos. En el ámbito de la reutilización Pronk (2019) expone sobre los beneficios que implica la reutilización de datos de investigación, dentro de los que sobresale un eficiente aumento de la productividad por parte de comunidades científicas, también al ser los datos de investigación recursos que están disponibles resultan útiles para la sociedad en general, ya que se ahorra tiempo que ocuparía llevar a cabo los procesos de recopilación y generación de datos, por lo tanto, permitirle a las distintas comunidades científicas en lugar de dedicarse a nuevas investigaciones apoyarse en los datos que están a disposición para reutilizar o también presentar otros provechos como servir en escenarios para etapas de enseñanza o aprendizaje.

De acuerdo con los hallazgos de Tang y Hu (2019) los servicios de gestión de datos de investigación (RDM, por sus siglas en inglés) tienen la tarea de organizar y llevar a cabo una serie de actividades relacionadas con el ciclo de vida de los datos de investigación, incluyendo el diseño, creación, reutilización y seguridad de los datos. Los servicios que ofrecen los RDM incluyen la planificación, intercambio, difusión, conservación, acceso y descubrimiento de datos, así como la gestión de metadatos y la visualización de datos. Además, el estudio incluye una encuesta internacional sobre los servicios RDM, en la que el 77,2% de los participantes indicaron que utilizan repositorios de datos como herramientas para la gestión de datos. Además, se hallan ciertos obstáculos y se vislumbraron posibles roles futuros para los servicios de gestión de datos de investigación (RDM).

También es importante lo expuesto en (Grasse et al., 2018) donde se aclara que un enfoque único en las soluciones técnicas como las de un repositorio es insuficiente para promover las actividades de RDM y aplicar los procedimientos de RDM en todo el ciclo de vida de la investigación, es fundamental el enfoque en las personas capacitadas.

En Adriaan et al. (2020) se subraya sobre la gestión de datos de investigación (RMD) lo cual tiene como fin administrar y mantener la organización de los datos, en busca de que el proceso de investigación resulte eficiente y genere facilidades en términos de cooperación con otros investigadores a través de favorecer la disponibilidad de los datos, el acceso y reutilización cuyas características se valen de los principios FAIR como eje central para lo que este ligado a gestión de datos de investigación, se menciona también la trascendencia que puede tener implementar algunas políticas las cuales son fundamentales siempre y cuando se mantengan cercanas a la práctica investigativa, dado que existe una diferencia entre lo que son las políticas y lo que es realmente la práctica investigativa, por ejemplo, en términos de propiedad intelectual.

En Xu et al. (2022), se muestran los efectos de una instrucción diseñada con base en el ciclo de vida de los datos de investigación y en el conocimiento de RDM en línea puede ser una alternativa efectiva para capacitar personas en estos conocimientos.

Schöpfel et al. (2018) sostiene que los financiadores y la Comisión Europea promueven a que las comunidades científicas se comprometan más con la gestión de datos de investigación (RDM) y también a sumar esfuerzos a la apertura de los datos de investigación. En este trabajo se describe que los principios de datos FAIR se aceptaron como principios rectores para garantizar que los datos sean fácilmente encontrados, accesibles, interoperables y reutilizables. Los principios FAIR describen cuatro dimensiones para evaluar diferentes niveles de FAIR y son vitales en RDM, incluso se destaca que para abordar adecuadamente RDM es esencial considerar las necesidades, demandas y comportamientos de las comunidades científicas, además de proporcionar orientación, capacitación y apoyo. Es importante considerar que, en vez de enfatizar la acción de compartir, es fundamental que la política de datos se enfoque en promover las buenas prácticas en RDM.

En Hettne et al. (2020), la Universidad de Leiden (LU) lanzó una reglamentación para ayudar a los investigadores cumpliendo con requisitos RDM, exigidos por patrocinadores y agentes para mejorar la transparencia e integridad de la investigación. Esta reglamentación tuvo como pilares, trabajos de organizaciones internacionales como: *GO FAIR*, *Research Data Alliance*, *CODATA*, las cuales son organizaciones que plantean y fomentan buenas prácticas en implementación de los principios FAIR. Dentro de la reglamentación se propone para lo que



corresponde a la etapa inicial, se deba empezar por un plan de gestión de datos (DMP) para luego continuar con la recolección de los datos. Luego se propone que, para la recolección de datos, se garantice que los datos se conserven de modo que, en la etapa siguiente, se garantice la integridad y disponibilidad de los datos mediante documentación y metadatos, así como garantizar la confidencialidad de datos. Se sugiere que para garantizar el acceso y sostenibilidad de los datos es una buena alternativa el uso de un repositorio con *CoreTrustSeal*. Adicionalmente, referente al tiempo en relación con la preservación de los datos que sugieren que sea por un rango de 10 años.

Según lo expuesto en Mushi et al. (2020) la falta de una política de RDM por parte de una institución no implica que una biblioteca deje de desarrollar habilidades de RDM o de brindar servicios de RDM, por el contrario, se deben desarrollar e implementar sistemáticamente componentes y servicios de RDM teniendo en cuenta lo que se ha demostrado en la literatura como buenas prácticas y así crear conciencia dentro de la organización y con el tiempo, garantizar que los datos de investigación de la universidad sean accesibles para la comunidad en general.

### **2.3.1.3 Desafíos de RDM**

En Lefebvre et al. (2020) se muestran algunos desafíos organizacionales y tecnológicos que ocurren durante la fase de planificación de la gestión de datos de investigación, desafíos que se deben abordar según los autores a partir de criterios de calidad que se basan principalmente en principios FAIR, también recalcan gran importancia en la reutilización, el uso de identificadores persistentes y el manejo de estándares de metadatos relevantes como Dublin Core o Data Cite.

Al analizar la documentación y publicación de datos de investigación surgen interrogantes en torno a la falta de incentivos para los autores ante el gran esfuerzo que puede conllevar este proceso como se menciona en Bloemers y Montesanti (2020). Según Cousijn et al. (2019), es importante definir métricas de datos con las que los datos de investigación pueden ser parte de las evaluaciones de calidad para proyectos de investigación y así recompensar a los autores y fomentar la publicación de datos de investigación.

Otro aspecto encontrado como reto, es la falta de políticas relacionadas con disponibilidad de datos, por ejemplo en Graf et al. (2020) se analizan declaraciones de disponibilidad de datos enviadas a 176 revistas entre 2013 y 2019, se aprecia que el aumento de declaraciones de disponibilidad de datos coincide con el lanzamiento de políticas por parte de las revistas que piden estas declaraciones como requisitos, esto ayuda a mejorar la ausencia de declaraciones de disponibilidad de los datos, un caso similar se encuentra en Jackson (2021), donde se recomienda la definición de políticas de manera clara teniendo en cuenta tipo de datos y cómo se deben aplicar; en relación a licencias se detalla de la búsqueda de un acceso libre y sin restricciones o dado el caso atribuir crédito al autor original,

puntualmente con la licencia (CC0) o (CC-BY), además de la utilización de repositorios generalistas como Dryad y figshare. De igual manera, en relación con las licencias y derechos de autor Subirats-coll et al. (2022) mencionan el uso de la licencia CC-BY.

### **2.3.2 Metadatos**

Teniendo claro que los datos de investigación son información generada o recolectada a lo largo de un proceso investigativo, es importante definir que esta información requiere ser descrita, como se menciona en Korkeamäki et al. (2022) “Las características de los datos son las propiedades y aspectos que describen los datos.”, se trata por tanto, mediante los metadatos brindar un contexto sobre los datos generados durante el proceso de investigación, por consiguiente los metadatos resultan ser muy útiles para describir otros datos, que para el contexto de nuestro trabajo serían los datos de investigación. Así se evidencia que los metadatos se convierten en un mecanismo esencial para el enriquecimiento de los datos de investigación y pueden ser FAIR. Hacia este mismo objetivo apunta Schröder et al. (2022) cuando afirma que “los metadatos son importantes para proporcionar un contexto adicional, lo que permite que otras personas (incluido el yo futuro) comprendan el proceso de investigación y los datos resultantes”. Asimismo, en Rumble et al. (2019) agregan que los metadatos son trascendentes debido a que “los metadatos son los datos y la información que permiten al usuario comprender la generación, el significado y la calidad de los datos”. Los metadatos son esenciales cuando se trata de facilitar al usuario la descripción de los datos (Calhoun et al. , 2019).

Los metadatos según Ganske et al. (2020) se entienden como datos que describen aspectos característicos de otros datos en particular, los metadatos se encargan de brindar la información necesaria de manera detallada acerca de cómo fueron generados los datos, un ejemplo de metadato puede ser la fecha de creación.

El uso de metadatos es esencial para los procesos de publicación y preservación los datos de investigación y como se ha venido mencionando anteriormente, los metadatos se convierten en la información relevante que brinda el contexto desde cuando fueron creados los datos de investigación (Töwe & Barillari, 2020).

Dentro de las ventajas que representan los metadatos en datos de investigación se puede subrayar algunas como la de permitirle al investigador que desea posiblemente reutilizar los datos de investigación poder determinar si esos datos son aptos o no para los escenarios de sus aplicaciones. Sumado a esto, los metadatos son una fuente con la cual el usuario interesado en los datos de investigación determina la calidad de los mismos, los metadatos hacen que los datos sean detectables, accesibles y reutilizables (Hashim, 2019), (Masinde et al., 2021). En ese mismo sentido, Enescu et al. (2018) habla sobre los propósitos de los metadatos y se sostiene que los metadatos se convierten en documentación de

los datos de investigación almacenados en un repositorio y esto permite tanto la facilidad de búsqueda como la accesibilidad de dichos datos para los usuarios interesados, además, de puntualizar que la calidad de estos metadatos va a facilitar la encontrabilidad de los datos.

La importancia de los metadatos la subraya Hansson y Dahlgren (2022) cuando manifiesta que gracias a los metadatos cualquier usuario interesado en los datos de investigación pueda tener una adecuada descripción acerca del conjunto de datos de investigación, así como un mejor conocimiento sobre cómo fueron recolectados los datos y con ello facilitar tareas que requieran la reutilización de los datos de investigación para el contexto que se necesite. Quarati y Raffaghelli (2022) en su trabajo recalca la importancia de los metadatos cuando menciona que una buena calidad de metadatos permite que los usuarios o investigadores puedan explotar eficientemente los recursos, pues contarían con la posibilidad de consultar fácilmente los datos para efectos de replicar o reutilizar dichos recursos de investigación.

Durante esta revisión sistemática se puede evidenciar la gran variedad de datos de investigación que se pueden producir dependiendo de la disciplina en particular, por esta misma razón los metadatos utilizados para describir los datos de investigación son muy diferentes, como se puede apreciar en Brandt et al. (2022) donde la disciplina es ciencia de los materiales y los metadatos utilizados son por ejemplo temperaturas, tiempos, materiales, entre otros, a diferencia de Korkeamäki et al. (2022), donde la disciplina es estudio de medios, donde los metadatos son por ejemplo nombre, direcciones, estadísticas, opiniones.

Según lo anterior, ante las diferentes necesidades que se pueden tener dependiendo de la disciplina en particular, se evidencia la necesidad de manejar un lenguaje en común en cuanto a metadatos de datos de investigación, esto para lograr según los principios FAIR la interoperabilidad. En este sentido, algunos autores proponen que para que exista un vocabulario común en cuanto a metadatos y para efectos de apoyar a los generadores de producciones científicas se opte por esquemas de metadatos flexibles dependiendo de la disciplina y en esa flexibilidad se recurra al uso de plantillas a fin de apuntar hacia un estándar (Brandt et al., 2021).

También autores como Hrynaszkiewicz et al. (2020) mencionan que para lograr tener un lenguaje en común se deben seguir los lineamientos de un estándar de metadatos de investigación que es “una forma común e interoperable de representar, etiquetar o estructurar datos y metadatos”. Teniendo en cuenta que un estándar responde a unas necesidades puntuales, se podría usar un estándar en particular que se ajuste a las necesidades del contexto y tipo de datos de investigación que se van a tratar en el prototipo planteado en el presente trabajo.

### 2.3.3 Estándares de metadatos

A partir de los artículos revisados, se ha encontrado inicialmente que un estándar de metadatos es “una forma común e interoperable de representar, etiquetar o estructurar datos y metadatos” (Hrynaszkiewicz et al. 2020). Por consiguiente, los estándares desempeñan un papel importante dentro de un repositorio de datos de investigación a raíz de que fomenta establecer un vocabulario común y por consiguiente facilita aspectos como la interoperabilidad, facilitar la legibilidad de las máquinas y en el transcurso de la revisión de diferentes artículos coinciden en que uno de los estándares ampliamente adoptado es Dublín Core (Jejkal et al., 2022). Además, de la interoperabilidad y encaminara que los datos de investigación sirvan a otros investigadores, los estándares son un pilar fundamental en la gestión de datos de investigación, considerando que en todo proceso investigativo es relevante el uso de estándares, aplicados desde la toma de datos hasta la publicación de estos datos, ya que puede fomentar la disponibilidad y la comprensión de los datos para facilitar la reutilización de dichos datos. Según Schirrwagen et al. (2019), un estándar de metadatos apoya y promueve la transparencia de la investigación y en ese sentido se menciona que un estándar adicional es el *Data Documentation Initiative* (DDI). De igual modo Hrynaszkiewicz et al. (2020) en su estudio sostiene que el acogerse a formatos estándares y comunes de metadatos ayuda significativamente en procesos tanto de descubrimiento como de una eficiente reutilización de los datos de investigación.

Otros autores como Latif et al. (2021) sostienen que algunos estándares adicionales de metadatos son DataCite, también se menciona (OAI-PMH) el cual es un protocolo para la recolección de metadatos y que permite servir más de un esquema de metadatos, DCAT, DDI. Igualmente, en Klump et al. (2021) nombran modelos de metadatos específicos de dominio tales como Darwin Core (DwC), DataCite, Dublín Core. Mientras que, Thessen et al. (2019) en su trabajo destaca el estándar de datos para la biodiversidad Darwin Core, COPDESS estándares para publicación de datos en ciencias de la tierra y DataCite como estándar de publicación de datos.

El vocabulario de Schema.org, que abarca entidades, relaciones y acciones, se presenta como una solución flexible para asegurar la interoperabilidad en diferentes codificaciones, como RDFa, Microdata y JSON-LD (Schema.org, n.d.). Esta versatilidad lo convierte en una alternativa en el ámbito de los repositorios, respaldada por estudios que destacan su eficacia en garantizar la interoperabilidad (Gaignard et al., 2023; Tsueng et al., 2023).

En Medina-Smith et al. (2021) se plantea un esquema de metadatos codificado mediante XML, el cual se basó en esquemas y vocabularios existentes, como los que se mencionaron anteriormente, Dublín Core y DataCite y también se revisa vocabulario relacionado con la ciencia de los materiales que es la disciplina puntual del trabajo realizado.

Otro trabajo donde se menciona un estándar de metadatos en particular es en Wasner et al. (2018) donde se recolectan datos de unas encuestas que brindarán información sobre el desempeño del sistema educativo en Suiza, se expone que aspectos a mejorar es pasar de sus propios estándares patentados a estándares de metadatos adecuados y estandarizados como por ejemplo DDI que se ajusta a sus necesidades particulares.

En Stocker et al. (2020) se destaca la importancia de identificador persistente para los instrumentos relacionados con datos de investigación, el cual se encuentra alineado con los principios FAIR, en este trabajo se plantean metadatos FAIR sobre instrumentos y se analizan estándares de metadatos como ePIC y DataCite, con los cuales se adopta e implementa el esquema propuesto por los autores ya que resuelven las necesidades propuestas para la temática del artículo relacionada con identificadores persistentes y metadatos de instrumentos.

El hecho de adoptar un estándar de metadatos es importante para asegurar un almacenamiento estructurado de la información y el beneficio que surge es poder realizar un análisis apropiado de los datos de investigación sin inconveniente.

En Hansson y Dahlgren (2022) su revisión sobre repositorios abiertos destacados como ResearchGate Data, Dataverse de Harvard, Figshare, Zenodo tienen como común denominador los siguientes estándares de metadatos: DataCite Metadata Schema, Dublin Core. Enescu et al. (2018) coincide en estándares como Dublin Core, dataCite o ISO 19139 (ISO, 2007).

También en Nitecki y Alter (2021) se muestra la colaboración que se da entre un proveedor de tecnología y la biblioteca de una de las universidades más antiguas y reconocidas de Escocia, la Universidad de Glasgow, en la descripción de esta colaboración, expone la importancia de los estándares de metadatos y manejo de repositorios de datos de investigación de confianza en general para la adopción de los principios FAIR.

Iglezakis (2020) presenta un modelo personalizado de metadatos orientado a ingeniería computacional denominado EngMeta el cual se basa en los estándares de DataCite, Codemeta, ExptML y PREMIS, este es un modelo para la extracción automática de metadatos de archivos de simulación y funciona como núcleo descriptivo en un repositorio institucional. Además, dentro de este trabajo se mencionan otros estándares como CERA, el cual está enfocado para datos con temas orientados al clima.

Se puede apreciar que los estándares de metadatos tienen utilidades apropiadas para diferentes necesidades, cada estándar de metadatos puede ser útil para un contexto en particular y ante la necesidad de tomar características de diferentes estándares de metadatos para cubrir diferentes necesidades, en Miksa et al. (2021)

se describe lo que se denomina perfil de aplicación y se establece uno para DMP procesables por máquina, teniendo en cuenta los estándares Dublin Core, Dublin Core Metadata Terms, - *Friend of a Friend* (FOAF), *DataCite Metadata Schema*, CERIF, COAR, Norma ISO 6391-1 y Norma ISO 4217.

A continuación, se definen los principales estándares encontrados en esta revisión sistemática.

- Dublin Core: “un conjunto de quince elementos (propiedades) "básicos" para describir recursos” (Dublin Core, n.d.)
- Datacite: “Esquema de metadatos de DataCite para la publicación y citación de datos de investigación y otros resultados de investigación” (DataCite, n.d.)
- DDI: “Un conjunto gratuito de estándares/productos originados por profesionales de datos (dentro de las ciencias sociales, económicas y del comportamiento)” (DDI Alliance, n.d.)

#### **2.3.4 Repositorios de datos de investigación**

Los repositorios de datos de investigación son un mecanismo que sirve a diferentes comunidades académicas y disciplinarias, es un intermediario, son entendidos como mediador entre productores y usuarios que consumen datos de investigación científica, al mismo tiempo los repositorios de datos de investigación benefician también los intereses de los patrocinadores y principalmente a las comunidades científicas, son un medio con la capacidad de preservar el valor de los datos a largo plazo según Štebe et al. (2020). En este mismo sentido, en Mushi et al. (2020) destacan que los repositorios de datos de investigación son una buena práctica de RDM dado que son una herramienta importante y oportuna dentro de la infraestructura de la investigación, en especial cuando se trata de preservar los datos valiosos que se han elegido para un uso a largo plazo, así mismo Koltay (2020) expone que para garantizar el uso y la conservación a largo plazo de los datos, las organizaciones de financiación y los editores deben incentivar a los investigadores para que depositen sus datos en repositorios certificados y acreditados según los estándares apropiados, ya que esto es clave para garantizar la calidad de los datos

Hansson y Dahlgren (2022) en su revisión encuentran que los repositorios de datos no solo le permiten al investigador publicar sus datos de investigación, sino que también potencializa que los investigadores puedan trabajar de manera colaborativa, con el propósito de que exista una transparencia y mayor visibilidad de las producciones científicas en donde los datos de investigación recolectados puedan ser reutilizables y escalables en el futuro. Precisamente al hablar de reutilización, en Van De Sandt et al. (2019), se busca una definición de reutilización de datos de investigación y se denota la situación de reutilización cuando un

individuo publica los datos de investigación en un repositorio, otra persona toma estos datos publicados y los utiliza en su nueva investigación.

En Aleixandre-benavent et al. (2020), a partir de una encuesta realizada a 1.063 investigadores en varias disciplinas y con el fin de obtener información sobre los hábitos y experiencias de los investigadores en España con relación a la gestión y el intercambio de datos brutos de investigación, se destaca que el 54.9% de los encuestados no cuenta con un plan de gestión de datos. Por otra parte, para lo relacionado a preservar los datos, la encuesta arroja que el 56.6% de los encuestados consideran que un repositorio es un recurso adecuado, particularmente junto a otros estudios se encuentra que el principal medio para acceder y localizar datos de otros investigadores es a través del contacto con colegas. En cuanto al intercambio de datos, se identifican algunas dificultades, como aspectos éticos, la pérdida de autoría, el uso inadecuado de los datos (por ejemplo, una interpretación errónea) y otro aspecto que tiene que ver con asuntos legales. En este trabajo se menciona la importancia del intercambio de datos, ya que permite acelerar y generar conocimiento, para aprovechar los datos existentes; además, enfatizar que muchas entidades financieras apoyan este tipo de iniciativas de intercambio de datos con el objetivo de recibir un mayor retorno a las grandes inversiones realizadas.

Downs (2021) resalta que en muchos proyectos de investigación científica se hacen grandes inversiones para la obtención de los datos de investigación científica y en este sentido los repositorios se destacan como una plataforma que puede llegar a diversas audiencias y con la cual se podría minimizar estos costos asociados. Al depositar los datos en un repositorio, estos datos se conservan y son accesibles para el uso continuo en nuevos proyectos de investigaciones científicas y puedan ser reutilizados, conduciendo a un mayor valor científico a partir de nuevos descubrimientos y usarlos con nuevos propósitos fomentando de esta manera la actividad de la investigación, cuyos resultados generen productos innovadores, servicios y estudios que beneficien a la sociedad. Los datos de investigación aparte de requerir grandes inversiones en términos de costos, también, implican un esfuerzo considerable (Masinde et al., 2021).

En aspectos de visibilidad, reutilización, transparencia, respaldo, mejorar la reproducibilidad y descubrimiento de datos de investigación que puede brindar un repositorio de datos de investigación también se suman otros autores como Arend et al. (2020), Defranco y Voas (2021), Jiao y Li (2022), y Klump et al. (2021). Adicionalmente, Posavec et al. (2020) en su estudio resalta a los repositorios, como un sistema electrónico cuya función se basa en capturar, difundir, así como preservar los resultados intelectuales de una institución y miembros de tal comunidad y cuyos resultados fueron fruto de una investigación científica.

En Bossaller y Million (2022) se plantea el potencial de los repositorios para acelerar e impulsar la creación de nuevos conocimientos científicos con investigadores que crean conocimientos mediante la reutilización de datos archivados, así mismo, es fundamental destacar la gran relevancia de que los datos archivados sean abiertos y reutilizables. En palabras de Ishak et al. (2019) “los factores más influyentes en el desarrollo o diseño del repositorio de datos de investigación son la tecnología, los datos y metadatos y las políticas”.

Con los repositorios de datos de investigación se logra una eficiente preservación de los datos y sirven de canal para que quien realice investigaciones científicas pueda crear, intercambiar o reutilizar recursos como los datos de investigación, además se plantea que los repositorios permiten gestionar datos de investigación para lograr una mejor calidad de la información según Lefebvre (2023). Los repositorios de datos de investigación tienen como enfoque principal, proporcionar datos de investigación que sean accesibles y reproducibles (Brandt et al., 2021).

Se presentan algunos desafíos como lo es el tema de costo que puede implicar el almacenamiento de la información recopilada como consecuencia de la constante evolución y alto volumen de información y la continua cantidad de datos obtenidos para determinados campos, por lo que habría que considerar acordar tiempos de preservación, por otro lado, se resalta el proceso de certificación de los repositorios lo cual ayude a desarrollar confiabilidad en los repositorios, como por ejemplo, la certificación para repositorios *CoreTrustSeal* (Töwe & Barillari, 2020)(K. Gregory et al., 2018) (Brandt et al., 2021). En el aspecto de poner límites de tiempo para la preservación de la información, por ejemplo, el gobierno de China exigió que se mantenga disponible la información durante 20 años (Marlina et al., 2022). Otra adversidad está relacionada con diferentes aspectos, tales como cuestiones legales y éticas al tratar con datos abiertos, en ese sentido como demostración podemos encontrar 11 impedimentos legales y algunas medidas para contrarrestarlos, esto se presenta en Hashim (2019).

Los repositorios también se pueden considerar como una infraestructura que se usa principalmente para publicar datos, lo cual implica que sirven para apoyar a los investigadores en sus procesos de investigación, los datos de investigación también se almacenan y gestionan de forma estructurada y para posteriores análisis, rompiendo barreras y convirtiéndose en instrumentos centrales de gran apoyo para las comunidades científicas en temas referentes a intercambio de datos, los repositorios promueven la reproducibilidad; se hace hincapié en que existen algunos repositorios o sistemas de código abierto para uso interno de grupos de investigación que permiten personalizarse y dentro de los que se destaca CKAN, Dataverse, DSpace, Invenio (Brandt et al, 2021). Otros repositorios nombrados en otros trabajos como en Jiao y Li (2022), son re3data el cual es uno de los indexadores de repositorios más completos, al igual que el caso de *Open Science Framework* (OSF), Dryad, Zenodo, y Figshare.



En Hedeland (2020) se describe la infraestructura digital en un centro de datos de investigación para datos de investigación lingüística audiovisual, donde un repositorio denominado HZSK, es el componente central de la infraestructura, componente que fue desarrollado teniendo en cuenta los requisitos generales para los repositorios de datos de investigación como se definen en los principios FAIR, también cabe resaltar que en la infraestructura se hace uso de metadatos estandarizados como CMDI y Dublín Core, también en Lehmann et al. (2023), se propone un sistema de recopilación y administración de datos de investigación alineado con los principios FAIR, siendo uno de sus componentes principales un repositorio, se mencionan alternativas como Dataverse y Zenodo para almacenar varios datos de investigación en repositorios.

Con el propósito de obtener un panorama actualizado sobre las implementaciones de repositorios de datos, en Bhardwaj (2019) se realizó un análisis de contenido de los Repositorios de Datos de Investigación abiertos (RDR) en todo el mundo, este análisis fue construido a partir de datos obtenidos de r3data y brinda información relevante que debe considerarse en relación a los RDR analizados, en particular se define tipo de licencias utilizadas, donde se evidencia que se hace uso principalmente de la licencia *Copyright*, seguidos de 'CC0', 'Dominio público', *Open Data Common* (ODC), *Open Government License* (OGL), *Open Government License Canada* (OGLC) y Apache License 2.0, este estudio también brinda información de los estándares de metadatos más utilizados en los repositorios a nivel mundial, donde se destacan Dublín Core en primer lugar, seguido de ISO 19115, DDI, entre otros. También se brinda información de los softwares más utilizados para la implementación de los RDR, encontrándose en primer lugar Data Verse, seguido de MySQL, DSpace, CKAN, Fedora, EPrints, entre otras, incluso se indica cuáles son los países líderes en el establecimiento de repositorios de datos de investigación, donde se encuentra en primer lugar Estados Unidos, Alemania, Reino unido, Canadá y Francia, por último se puede apreciar cuáles son las temáticas de los repositorios más habituales siendo en orden: la biología, geociencias, medicamentos, ciencias sociales, físicas y humanidades.

Es de resaltar la relevancia de las licencias y derechos de autor en la temática de los datos abiertos, en Minamiyama et al. (2020) se expone la dificultad que puede generar políticas demasiado restrictivas y directrices poco claras sobre el intercambio de datos, se habla sobre el uso de herramientas estandarizadas, como las licencias *Creative Commons* (CC) y *Open Data Commons* (ODC) y otras licencias CC planteadas como licencias diseñadas para proporcionar condiciones de uso flexibles, teniendo en cuenta la ley de derechos de autor. Es importante enfatizar que, en la ciencia abierta, se busca promover el acceso libre y la reutilización del conocimiento científico, pero sin prescindir de los derechos de autor, algunas licencias como CC, incluyen condiciones que respetan el derecho de autor, pero permiten a los autores autorizar el uso de sus trabajos bajo ciertas condiciones.

En Poblet et al. (2018), se también se aborda sobre “la falta de licencias claras y la transparencia de los términos y condiciones de uso de los datos de investigación”, este mismo autor, por lo tanto, expone el uso de licencias *Creative Commons* para metadatos de investigación como alternativa para lograr que los datos de investigación estén conectados, sean detectables y reutilizables, se centra principalmente en la licencia CC0 y CC-BY, también se exponen protocolos y prácticas de metadatos de investigación como el Protocolo para la Recolección de Metadatos OAI-PMH y Dublín Core (DCMI).

En relación con el software utilizado para repositorios de datos de investigación también se encuentra que en Nie et al. (2021) se realizó una evaluación para la selección de un software que permita la implementación de un repositorio de datos de investigación en una universidad de Pekín, donde se probaron las herramientas *Dataverse*, *Data Conservancy*, *CKAN* y *DSpace*, también se describe algunas características particulares de estas herramientas, en relación a soporte, disciplinas y a su licenciamiento, al analizar estas características particularmente se destaca *DSpace* en relación a las necesidades del proyecto, principalmente por ser una herramienta multidisciplinaria, gratuita y que cuenta con soporte.

En Senagi y Tonnang (2022) se expone la definición de un sistema de gestión de datos de investigación donde en el componente de almacenamiento se seleccionó, instaló y configuró en la plataforma de código abierto que permite la gestión de datos y la publicación de conjuntos de datos en línea, *CKAN*, también se contempló el uso de *Dataverse*, se aprecia el uso de la licencia *Creative Commons* en su entorno de trabajo para facilitar el acceso a los datos y sus metadatos y cabe resaltar el uso de *GitHub* como plataforma para almacenar códigos fuente de varios software o fragmentos de código de software.

En Hasselbring et al. (2020) se habla sobre software de investigación utilizado en el descubrimiento científico o como objeto de investigación en sí mismo, también se expone el uso de herramientas como *Github* para alojar y compartir proyectos de software, este tipo de datos también debe seguir los principios FAIR y se incita a fomentar la citación de software para vincular el software con publicaciones, conjuntos de datos, métodos y otros objetos de investigación, también se menciona que repositorios como *Zenodo* pueden servir para alojar instantáneas de software y se recomienda manejar las licencias de software de código abierto establecidas.

A continuación, se realiza la definición y descripción de algunos repositorios relevantes que han surgido en el ámbito de la gestión de datos de investigación y que fueron hallados en la revisión del conjunto de artículos primarios seleccionados.

- *Dataverse.org* (2023) define a *Dataverse* como “Software de repositorio de datos de investigación de código abierto”.

- Dryad es "una plataforma de publicación de datos abiertos y una comunidad comprometida con la disponibilidad abierta y la reutilización rutinaria de todos los datos de investigación". Inicialmente, Dryad fue creado "para seleccionar y publicar abiertamente una gran variedad de tipos de datos, como los asociados con la investigación en ecología y biología evolutiva". Importante tener en cuenta que distintas comunidades como instituciones académicas hacen inversiones para "la conservación, el intercambio abierto y la reutilización de datos de investigación en todos los campos". Además, "Dryad es completamente de código abierto" de acuerdo con Datadryad.org (2023).
- FEDORA (Flexible Extensible Digital *Object Repository Architecture*) es "un sistema de repositorio robusto, modular y de código abierto para la gestión y difusión de contenido digital. Es especialmente adecuado para bibliotecas y archivos digitales, tanto para el acceso como para la conservación. También se utiliza para brindar acceso especializado a colecciones digitales muy grandes y complejas de materiales históricos y culturales, así como a datos científicos". (fedora.lyrasis.org, 2023)
- Figshare es definido como "es un repositorio donde los usuarios pueden hacer que todos los resultados de sus investigaciones estén disponibles de una manera citable, compartible y detectable" y que es "adecuado para todos los tipos de contenido, no solo para datos, implementamos dominios personalizados". Es de destacar que Figshare trabaja con "una variedad de instituciones, editores, financiadores y conferencias para crear un entorno más eficiente para la investigación" (figshare, 2023).
- Zenodo es "un repositorio de datos de investigación de difusión abierta para la preservación y puesta a disposición de contenido de investigación, educativo e informativo. El acceso al contenido de Zenodo está abierto a todos, solo para fines no militares". Se destaca que "El código de Zenodo es en sí mismo de código abierto y se basa en la biblioteca digital Invenio, que también es de código abierto", al igual que "Zenodo rápidamente comenzó a recibir investigaciones de todo el mundo y de todas las disciplinas" (zenodo.org, 2023).

Es oportuno precisar que CERN (*European Council for Nuclear Research*) se encarga de investigar "la estructura fundamental de las partículas que componen todo lo que nos rodea. Lo hacemos utilizando los instrumentos científicos más grandes y complejos del mundo" (cern, 2023).

- De acuerdo con inveniosoftware.org (2023), Invenio es un proyecto de código abierto que fue creado originalmente por el CERN. En particular, el producto InvenioRDM sobresale por ser "una plataforma de gestión de

repositorios/documentos” expresado también en (inveniosoftware.org, 2023)

- DSpace se centra en repositorios institucionales además de ser un software ágil, flexible, adaptable, de código abierto, que “satisface una necesidad específica como sistema de archivos digitales, centrado en el almacenamiento a largo plazo, el acceso y la preservación del contenido digital, lo que convierte a DSpace en el software de elección para organizaciones académicas, sin fines de lucro y comerciales que crean repositorios digitales abiertos” (DSpace.lyrasis.org, 2023).
- CKAN el cual es “un DMS (sistema de gestión de datos) de código abierto para impulsar centros de datos y portales de datos. CKAN facilita la publicación, el intercambio y el uso de datos. Impulsa cientos de portales de datos en todo el mundo” (ckan.org, 2023)

### **2.3.5 Sistemas de información de investigación actual (CRIS)**

Para Schöpfel et al. (2017) los sistemas CRIS son necesarios para manejar la gran complejidad de los datos científicos y están diseñados para almacenar y administrar datos de investigación de una organización o institución. Además de recopilar y analizar información, estos sistemas también proporcionan acceso y difunden información de investigación relevante. Se considera que los CRIS son una opción para mejorar los servicios de biblioteca en la gestión de datos de investigación, con su enfoque principal en el enlace y almacenamiento de datos. Los CRIS pueden mejorar el control de procedencia y el flujo de trabajo de los datos mediante el uso de identificadores, un lenguaje común, una semántica rica y enlaces a publicaciones relevantes. Además, de poder mejorar la calidad y la eficiencia de los datos.

Siguiendo los resultados de Rousi, (2022) los CRIS son sistemas de información de investigación que se utilizan para recopilar información sobre publicaciones científicas y perfiles de investigadores en organizaciones de investigación. Además, los CRIS de última generación incluyen métricas de citas integradas y multimedia para mejorar su funcionalidad. Aunque los estudios previos se han centrado en las funcionalidades relacionadas con la gestión de datos de investigación, aún no se ha aprovechado todo el potencial de los CRIS para obtener información sobre el intercambio de datos entre los investigadores.

En Sivertsen (2019) se afirma que los CRIS son cada vez más utilizados para estandarizar, facilitar y administrar la documentación y comunicación en la investigación, o, en otras palabras, para administrar actividades propias de la investigación científica. Además, con una amplia cobertura y datos de calidad estandarizados e integrales, los CRIS también pueden usarse como fuentes de datos para estudios de investigación y para tomar decisiones a partir de monitorear las actividades y resultados de la investigación. Para avanzar en su desarrollo con

finés de investigación, es necesario que los sistemas CRIS sean interoperables y comparables entre instituciones y países.

Para que un CRIS sea eficaz, según Dvořák et al. (2019) es importante que colabore con otros sistemas tanto dentro como fuera de la organización, incluyendo el intercambio de información. De hecho, un CRIS funciona como un sistema centralizado que mantiene una relación y conexión entre diferentes fuentes de información. Así, el CRIS puede recopilar y mantener datos actualizados y precisos y mantener información precisa y actualizada de múltiples fuentes. La capacidad de integración e interoperabilidad de los sistemas CRIS ha sido reconocida como uno de sus principales beneficios.

Según Fabre y Egret (2020) es conveniente utilizar un modelo de datos estándar e interoperable, especialmente el formato estándar internacional recomendado por la Comisión Europea para el sistema de gestión de información de investigación, Formato Común Europeo de Información sobre Investigación (CERIF), estos autores, señalan que el CERIF es un modelo estándar y genérico que se utiliza para organizar y compartir información sobre investigaciones, su objetivo principal es estandarizar la organización e intercambio de información sobre investigaciones y su dominio, en los niveles conceptual, lógico y físico, para lograr un acceso homogéneo a sistemas de datos heterogéneos y un formato de intercambio de datos definido. Además, se destaca que el modelo CERIF fomenta la integración y el intercambio de datos a través de la estandarización y funciona como nivel de interoperabilidad entre la infraestructura digital y los datos de investigación. Este modelo incluye información sobre personas, organizaciones, proyectos y financiación, así como cualquier otro elemento relacionado con el proceso de investigación, y tiene tres elementos centrales interconectados: personas, organizaciones y proyectos.

Los CRIS, según lo expuesto en Farinelli y Zigoni (2022), permiten la gestión de los aspectos relacionados con la información de investigación, los mismos autores muestran la integración de estos sistemas con herramientas de gestión de datos de investigación especializadas como repositorios, este artículo describe la adopción de estos sistemas en una institución, definiendo la importancia de la interoperabilidad lograda a través de la utilización de estándares de metadatos como CERIF, Dublin Core u OpenAIRE, también se expone la importancia de la integración entre CRIS y los repositorios institucionales para el archivo de datos, pues se garantiza la accesibilidad y la preservación a largo plazo de los datos de investigación. También en Jetten et al., (2019) se destaca que los CRIS se encuentran respaldados por CERIF, cubren todos los aspectos de la información de investigación, involucra aspectos como registro, financiación, DMP, investigadores, resultados de investigación, datos generados, metadatos de datos generados, entre otros, en este artículo también se menciona que el CRIS comercial de mayor difusión internacional es PURE con más de 200 implementaciones en el mundo y

que el CRIS no comercial de mayor difusión internacional es DSpace-CRIS, con más de 100 implementaciones. PURE y DSpace-CRIS pueden almacenar colecciones de datos internamente gracias a las funcionalidades de repositorio integradas en CRIS, pero también están explorando la conexión a repositorios de datos externos existentes. En Zervas et al. (2019), se describe una transformación del repositorio institucional denominado KTISIS en un CRIS, esta transformación se implementó con el software DSpace-CRIS, en este trabajo se muestra también la relación directa entre repositorios institucionales y CRIS, definiendo que CRIS para gestión de datos sobre todos los aspectos de las actividades de investigación como proyectos, financiación, resultados y tecnologías y los IR para recopilar y exponer de manera gratuita los resultados de la investigación institucionales, esta relación en la transformación descrita en el trabajo se denota el describir que pasaron de una solución de repositorio institucional con DSPACE a un sistema CRIS usando DSpace-CRIS, herramienta que según estos autores es una “herramienta eficaz para recopilar y administrar información de investigación, como perfiles de investigadores, páginas de departamentos, subvenciones para proyectos, resultados de investigación, métricas, informes y estadísticas”, también es importante mencionar que su modelo de datos admite el uso de identificadores persistentes para crear relaciones navegables entre entidades, también se menciona el manejo de licencias CC.

En Biesenbender et al. (2019) se presenta la relevancia de los CRIS en la evaluación de resultados de instituciones en Italia, los Países Bajos y Alemania, describen como estas evaluaciones son de suma importancia a la hora de la asignación de recursos, particularmente en Italia se menciona el sistema IRIS, el cual es una plataforma CRIS basada en DSpace-CRIS.

De hecho, al hablar de CRIS en un contexto latinoamericano, se encuentra el trabajo de Rosalina Vázquez (2022), en el cual se realiza un estudio de diagnóstico sobre el desarrollo de los sistemas CRIS en América Latina donde se hace una encuesta en 133 instituciones de 16 países latinoamericanos donde 31 instituciones respondieron que sí cuentan con un CRIS, este estudio nos arroja resultados muy concretos en cuanto a software utilizados, ya que se encontraron 48 instalaciones de CRIS en donde 48% son PURE, 31% VIVO y 20% DSpace-CRIS, dándose un equilibrio entre soluciones comerciales y no comerciales, también se indica que los CRIS están estrechamente vinculados con RI.

### **2.3.6 Consideraciones finales de la revisión sistemática**

En resumen, esta sección presenta las conclusiones de la revisión sistemática, definiendo los aspectos principales a considerar en el diseño del prototipo a proponer en este proyecto.

Los datos de investigación son resultados de estudios aceptados por la comunidad científica, los cuales permiten validar y respaldar resultados, son datos que se

recolectan para realizar análisis o cálculos. Pueden adoptar varios formatos como, por ejemplo, encuestas, entrevistas, archivos multimedia, simulaciones, algoritmos, gráficos, entre otros. Son fundamentales para otros investigadores y pueden ser multidisciplinarios. Es esencial seguir los principios FAIR para garantizar que los datos sean localizables, accesibles, interoperables o reutilizables. Se debe recalcar que la gestión adecuada de los datos de investigación abarca todo el ciclo de vida, desde la planificación de la generación de datos hasta su intercambio, publicación y conservación.

Los repositorios de datos de investigación de acuerdo con la revisión sistemática ofrecen múltiples beneficios, que incluyen facilitar futuras investigaciones, promover la reutilización de datos, aumentar la eficiencia al ahorrar tiempo y recursos, mejorar la transparencia y reducir la duplicación de esfuerzos. Su existencia es fundamental para el avance y desarrollo de la comunidad científica ya que se encargan de almacenar, organizar y garantizar la disponibilidad de información a largo plazo de investigación científica. También es importante destacar que son fundamentales los principios FAIR en la construcción de estos repositorios de datos de investigación. Además, se reconoce en las diferentes publicaciones que la gestión de datos de investigación es una parte fundamental del proceso de investigación científica, y los repositorios de datos son una herramienta clave en este proceso.

El uso de estándares de metadatos en la gestión de la información según lo encontrado en la revisión sistemática es fundamental para lograr características necesarias como interoperabilidad, accesibilidad, reutilización, entre otras. Se trata de un lenguaje definido, estructurado y ampliamente utilizado. Existe gran variedad de estándares de metadatos los cuales tienen utilidad dependiendo del contexto y la necesidad, incluso en la revisión sistemática se encontró el término perfil de aplicación, que trata sobre utilizar diferentes estándares, para lograr satisfacer diferentes necesidades que en ocasiones no se pueden cubrir solo con un estándar.

En relación a las licencias como herramientas legales para la definición del uso de obras o recursos digitales se definen en la revisión sistemática como una necesidad fundamental en la temática de datos de investigación, se argumenta en repetidas ocasiones que la definición de una licencia estandarizada brinda beneficios a la hora de lograr la característica de disponibilidad, teniendo en cuenta la atribución de créditos al autor original, es de tener en cuenta que existe gran variedad de licencias, las cuales tienen características puntuales, que pueden ser apropiadas para un contexto o necesidad.

Con respecto a los CRIS respaldados por el modelo CERIF, según lo encontrado en la revisión sistemática son sistemas que permiten administrar y gestionar información relacionada a la investigación científica desarrollada en una institución u organización, los CRIS abarcan la gestión de diferentes aspectos asociados a las actividades de un proceso de investigación científica, como resultados, financiación,

investigadores, entre otros. Estos sistemas trabajan de manera conjunta con otras herramientas como repositorios para lograr mayor efectividad en la gestión de la información.



### 3. PROPUESTA DE MODELO DE PUBLICACIÓN DE DATOS DE INVESTIGACIÓN PARA UN ÁMBITO UNIVERSITARIO

En esta sección se presenta una iniciativa para abordar las necesidades específicas de gestión de datos de investigación en la Universidad del Cauca, mediante un prototipo de software para un grupo de investigación, el cual se denominará INVESTUNICAUCA, este prototipo se presenta como una propuesta que integra un modelo de publicación, el cual busca establecer un repositorio dedicado a la preservación, acceso y difusión de los datos de investigación generados por la comunidad académica de la universidad en consonancia con la gestión de la investigación. A continuación, se detallan las características y componentes del prototipo, así como las metodologías y estándares que guiaron su desarrollo. Además, se proporciona una visión general de INVESTUNICAUCA y su contribución como infraestructura de gestión de datos de investigación institucional inexistente.

En la siguiente sección, se presentan los modelos de representación de la propuesta que integra un repositorio de datos y sistemas CRIS. Estos modelos constituyen la base conceptual que respalda la integración de DSpace y CRIS, proporcionando una visión estructurada de la gestión de datos de investigación en el contexto de un grupo de investigación de la Universidad del Cauca.

#### 3.1 Modelos de representación de datos

Es importante resaltar que el modelo CERIF (*Common European Research Information Format*) fue clave debido a que permite la interoperabilidad y la gestión de información relacionada con la investigación. Este modelo fue diseñado por EuroCRIS, CERIF proporciona un marco estructurado y armonizado para la descripción de entidades, relaciones y actividades en el ámbito de la investigación. En este contexto, CERIF emerge como una herramienta clave para potenciar la visibilidad, accesibilidad y reutilización de datos de investigación en entornos académicos e institucionales. En ese sentido, se continúa a mostrar la definición de las principales entidades de CERIF, esto se puede visualizar en la Tabla 6.

Tabla 6. Definición de entidades modelo CERIF

Entidad	Definición	Atributos	Relaciones
<b>Agente</b>	Cualquier cosa que tenga la capacidad de realizar actividades.	palabras clave: Lista< Cadena multilingüe > vida útil: rango de fechas contactos: Lista< Información_de_contacto >	Un agente puede ser el actor de cualquier número de actividades. Un agente puede recibir cualquier número de premios. Un agente puede ser el agente en cualquier número de posesiones de conocimientos y habilidades.

			Un agente puede Sea el agente en cualquier cantidad de experiencia, tiempo y esfuerzo. Un Agente puede ser el agente en cualquier cantidad de Experiencia, Tiempo y Esfuerzo.
<b>Persona</b>	Especialización de Agente, Un ser humano como individuo.	nombre: Nombre de la persona nombres alternativos: Lista < Nombre de persona >	Las relaciones de Agente
<b>Organisation Unit</b>	Grupo formal, orientado a objetivos, con miembros y una estructura definida, regido por reglas y procedimientos que guían el funcionamiento	nombre: Cadena multilingüe acrónimo: cadena descripción: Cadena multilingüe	Las relaciones de Agente
<b>Project</b>	Un esfuerzo temporal emprendido para crear un producto, servicio o resultado único.	nombre: Cadena multilingüe acrónimo: cadena	No registra
<b>Event</b>	Algo que sucede, especialmente algo importante	No registra	No registra
<b>Document</b>	Representación limitada de un conjunto de información diseñado con la capacidad (y normalmente la intención) de comunicar. Un documento puede manifestar información simbólica, esquemática o sensorial-representacional.	título: Cadena multilingüe fecha de publicación: Fecha en caso de que el documento haya pasado por publicación	Un documento tiene cualquier número de contribuciones: instancias de Contribución al Documento. Un documento puede demostrar cualquier número de posesiones de experiencia y habilidades. Un documento puede demostrar cualquier número de actividades.
<b>Resource</b>	Cualquier cosa que pueda usarse para o en el proceso de investigación.	No Registra	Un recurso puede ser producido por cualquier número de Declaraciones de Contribución. Un recurso puede ofrecerse en cualquier número de Ofertas de Recursos. Un recurso puede tener cualquier número de <i>Resource_Identifier</i> .
<b>Activity</b>	Todo lo que hace un Agente.	rango de fechas: Rango de fechas	Una actividad normalmente tiene actor: una instancia de

			<p>Agente. Para algunas subclases de Actividad, es posible que el Agente no sea conocido o divulgado (por ejemplo, para revisiones ciegas). Además, si el actor utiliza un seudónimo, se desconoce su identidad (y, a veces, incluso el tipo), en cuyo caso solo está disponible el nombre para mostrar.</p> <p>Una actividad tiene cualquier número de Declaraciones de Afiliación.</p> <p>Una actividad puede ser acreditada por cualquier número de Documentos.</p> <p>Una actividad puede estar cubierta por cualquier número de Fondos.</p>
--	--	--	--

*Fuente: Tomado del repositorio del repositorio oficial de CERIF-Core en GitHub (EuroCRIS, n.d.). Traducido al español por los autores.*

En la Tabla 6 se describe la definición, los atributos y las relaciones correspondientes a las principales entidades del modelo CERIF.

A partir de lo anterior, se analiza las entidades del modelo CERIF que se deben tener en cuenta según las necesidades en el contexto institucional de la Universidad del Cauca, específicamente en el Grupo de Investigación en Inteligencia Computacional (GICO).

En este contexto y en concordancia con la revisión realizada con respecto a la caracterización de la fase anterior, se procede a tomar como base la información que reposa en el sitio web del grupo GICO (<https://www.unicauca.edu.co/gico/>) para llevar a cabo la identificación de la información relevante relacionada, con el objetivo de determinar las entidades pertinentes del CERIF que resultan esenciales para el ámbito del grupo de investigación universitario en consideración.

Según las principales características del modelo CERIF tal como se detallan en (EuroCRIS, 2020), se concluye que las entidades clave del núcleo de CERIF que resultan pertinentes para las necesidades del prototipo son las siguientes:

- *Person*
- *Project*
- *Organisation Unit*
- *Event*
- *Publication*
- *Resource*
- *Activity*

Como se evidenció en la fase anterior, dentro del proceso de investigación es muy importante la gestión de datos de investigación, siendo, los repositorios de datos de investigación un elemento clave para ese escenario, principalmente porque garantizan la disponibilidad de la información científica a largo plazo. En ese sentido, de acuerdo con los hallazgos en la revisión sistemática, es fundamental que en el prototipo del repositorio se incorpore el concepto ampliamente reconocido y respaldado con respecto a la publicación de datos y metadatos de investigación que son los principios FAIR.

Cabe resaltar que, para ese propósito, CERIF es un facilitador de interoperabilidad en lo que corresponde a la información de la investigación en sistemas CRIS. Por lo tanto, el aprovechamiento de las entidades del CERIF, que se presentarán en la Tabla 7 resultan relevantes, para cumplir con los principios FAIR. De las entidades del modelo CERIF mencionadas anteriormente contribuyen a alcanzar dichos principios, como se detalla a continuación:

*Tabla 7. Entidades que contribuyen a alcanzar los principios FAIR.*

	<b>F</b>	<b>A</b>	<b>I</b>	<b>R</b>
Person	x	x	x	
Project	x		x	
Organisation Unit	x		x	
Event	x		x	
Publication	x	x	x	x

En la Tabla 7, se presentan las entidades CERIF que desempeñan un papel fundamental en la implementación de los principios FAIR. En esta Tabla, se destaca la entidad específica, así como la identificación clara de qué principio FAIR facilita en el contexto de la investigación.

Seguidamente se detalla cómo cada entidad, presentada en la Tabla 7, contribuye a la realización de los principios FAIR en el contexto de la investigación.

Las entidades de la Tabla 7 para hacer que los datos sean Findable (F):

- *Person*: Utilizar esta entidad para describir a los autores y colaboradores de la investigación, proporcionando metadatos como nombres, afiliaciones y enlaces a perfiles ORCID.
- *Project*: Utilizar esta entidad para describir proyectos de investigación, incluyendo información detallada sobre el título del proyecto, los investigadores involucrados y las palabras clave.
- *Organisation Unit*: Utilizar esta entidad para describir las unidades y departamentos de la institución, lo que facilita la búsqueda y la vinculación de la investigación con las unidades correspondientes.
- *Event*: Utilizar esta entidad para describir eventos de investigación, como conferencias y simposios, lo que permite a los investigadores encontrar oportunidades para presentar sus resultados.

Las entidades de la Tabla 7 para hacer que los datos sean Accessibles (A):

- *Publication*: Utilizar esta entidad para describir las publicaciones de investigación y proporcionar información sobre el acceso a las mismas, como si son de acceso abierto o si requieren una suscripción.
- *Person*: Utilizar esta entidad para definir los roles y permisos de los investigadores y personal de la institución en relación con los objetos de investigación.

En la Tabla 7 se destaca el principio de interoperabilidad (I) dado que con las entidades de CERIF que ahí se muestran se logra definir relaciones entre las diversas entidades, por ejemplo, asociar publicaciones con autores o proyectos con investigadores, lo que contribuye a la interoperabilidad de los datos de investigación. Además, el uso de identificadores persistentes, como URIs o Identificadores de Recursos Uniformes (por sus siglas en inglés, *Uniform Resource Identifiers*) asignados para cada ítem perteneciente a cada entidad, potencia la interoperabilidad, permitiendo utilizar los identificadores de manera consistente por diferentes sistemas y aplicaciones.

Las entidades de la Tabla 7 para hacer que los datos sean Reutilizables (R):

- *Publication*: Proporcionar metadatos detallados en esta entidad, como resúmenes y descripciones de los resultados de investigación, para que los datos sean más reutilizables por otros investigadores.
- *Result o Publication*: Proporciona información sobre métricas y evaluaciones de impacto relacionadas con la investigación, lo que puede ayudar a los investigadores a comprender mejor el valor y la relevancia de los datos.

Una vez realizada la definición de las entidades para tener en cuenta del modelo CERIF, en la Figura 7, se presenta un esquema que describe de manera conceptual la organización y la relación de los datos dentro del sistema DSpace-CRIS, evidenciando principalmente la relación que se presenta entre las entidades del modelo CERIF y el repositorio DSpace.

A continuación, se realiza el análisis de la organización de los datos en DSpace, esto se puede apreciar en el modelo disponible en la documentación del software (LYRASIS, 2011), el cual se presenta en la Figura 7.

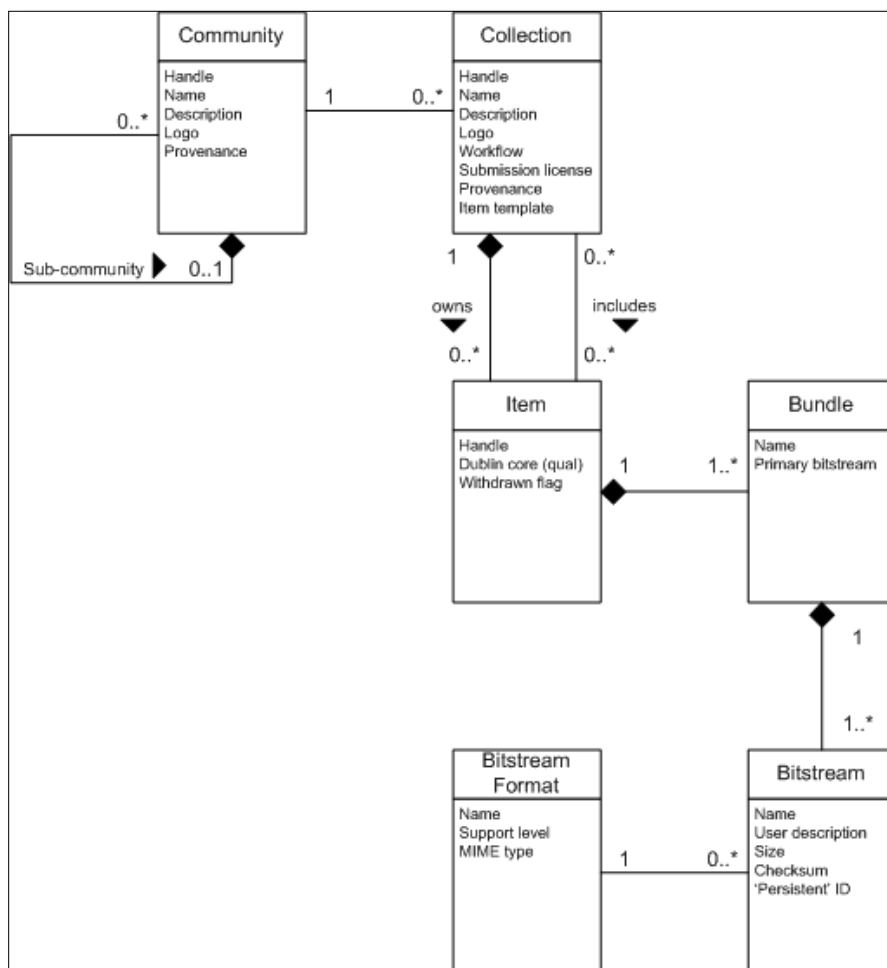


Figura 7. Modelo de datos de Dspace. Fuente: Tomado de: (LYRASIS, 2011)

En el modelo que se presenta en la Figura 7, se observa la organización de la información dentro de un DSpace convencional, donde de manera general se encuentran las comunidades, las cuales se pueden dividir en subcomunidades, estas comunidades son las que contienen las colecciones que son agrupaciones de contenido relacionado y las colecciones se componen de elementos o ítems y cada ítem tiene una y sólo una colección propietaria.

Los elementos se subdividen en paquetes de flujos de bits con nombre, los flujos de bits son archivos informáticos normales.

En el análisis de cómo se relaciona el modelo CERIF en la organización de la información en DSpace CRIS, teniendo en cuenta que, como característica principal, permite recopilar y gestionar datos de investigación e información propia de un sistema CRIS.

En el contexto del análisis detallado de la estructura organizativa del prototipo, se procede a presentar el modelo entidad-relación que vincula los módulos fundamentales de DSpace con los elementos CRIS.

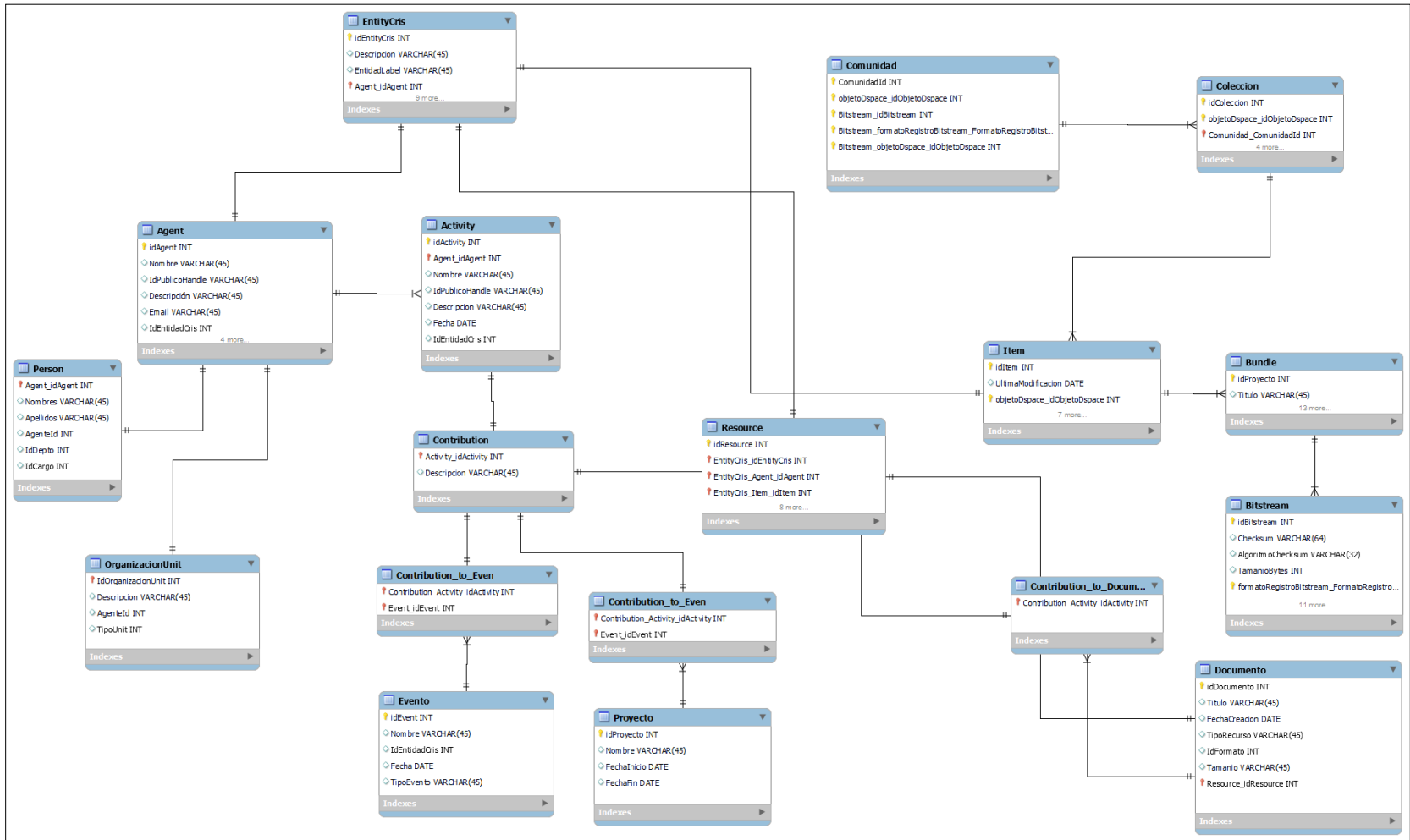


Figura 8. Modelo Conceptual de DSpace- CRIS para INVESTCAUCA.



Una vez definido el modelo conceptual en la Figura 8 se definirán los atributos que tendrá cada entidad o, en este caso, mediante qué metadatos pueden representar las entidades fundamentales para el prototipo según las necesidades del grupo de investigación al que va dirigido la presente propuesta de prototipo. Para lograr identificar los atributos necesarios se ha adoptado la estrategia de basarse en metadatos provenientes de Schema.org y OpenAIRE que utiliza el protocolo OAI-PMH v2.0 para recopilar metadatos y requiere que los metadatos estén codificados en el formato de metadatos Dublín Core.

Por consiguiente, mediante Schema.org y OpenAIRE se busca la representación de las entidades en el CRIS, ya que cuentan con la capacidad para proporcionar un marco estructurado y comprensible que abarca diversas áreas temáticas, incluyendo la investigación académica. Los tipos, elementos y propiedades definidos tanto en schema.org como OpenAIRE permiten una descripción detallada y estandarizada de las entidades relacionadas con la investigación, como proyectos, publicaciones, investigadores y otras categorías relevantes en el ámbito de investigación y académico. Schema.org y OpenAIRE como base para la representación de entidades en el CRIS contribuyen a establecer un marco robusto y compatible con estándares para la gestión efectiva de la información de investigación.

Después, se proporciona un análisis detallado de cada entidad, incluyendo los metadatos asociados a cada una. Este análisis tiene como objetivo proporcionar las características y atributos específicos de cada entidad, destacando su importancia y contribución al contexto general del prototipo desarrollado. La información presentada sirve como guía detallada para comprender las particularidades de cada entidad en el marco del sistema propuesto.

*Tabla 8. Atributos para la entidad persona*

<b>Atributo</b>	<b>Descripción</b>
Date	Fecha de nacimiento.
Email	Dirección de correo electrónico.
Telephone	Teléfono
Gender	Género de algo, normalmente una persona
Description	Una descripción del artículo
Identifier	La propiedad identificadora representa cualquier tipo de identificador para cualquier tipo de cosa, como ISBN, códigos GTIN, UUID, etc.
Image	Una imagen del artículo. Puede ser una URL o un ImageObject completamente descrito.
Url	URL del elemento.

La Tabla 8 proporciona una descripción concisa de los atributos asociados a la entidad Persona. En ella se detallan las características específicas y relevantes

relacionadas con la representación de individuos dentro del contexto del prototipo o sistema en consideración.

*Tabla 9. Atributos para la entidad Proyecto*

<b>Atributo</b>	<b>Descripción</b>
Name	Nombre del proyecto
DOI	Identificador de objeto digital
URL	Ruta del ítem
Type	Tipo de proyecto.
Person	Creador del proyecto.
Keyword	Palabra clave.
Date	Fecha.
Product ID	Identificador del producto.

La Tabla 9 presenta de manera concisa los atributos asociados a la entidad Proyecto. En esta Tabla se describen las características particulares y relevantes que definen y representan la información vinculada a proyectos en el contexto del prototipo.

*Tabla 10. Atributos para la entidad Organización*

<b>Atributo</b>	<b>Descripción</b>
Name	Nombre de la organización.
Parent organization	Suborganización
Identifier	Identificador de la organización.
Description	Descripción de la organización.
Type	Tipo de organización.

La Tabla 10 detalla de manera resumida los atributos relacionados con la entidad Organización. En este contexto, se describen las características específicas y relevantes que definen la representación de las organizaciones dentro del prototipo o sistema.

*Tabla 11. Atributos para la entidad Evento*

<b>Atributo</b>	<b>Descripción</b>
Name	Nombre del evento.
Type	Tipo de evento.
Start date	Fecha de nacimiento.
End date	Fecha final.
Description	Descripción del evento.

La Tabla 11 describe los atributos asociados a la entidad Evento. En este contexto, se detallan de manera concisa las características específicas y relevantes que definen la representación de eventos.

*Tabla 12. Atributos para la entidad Patrocinio*

<b>Atributo</b>	<b>Descripción</b>
Name	Nombre del patrocinio.
Amount	Monto del patrocinio.
Type	Tipo de patrocinio.
Description	Descripción del patrocinio.

La Tabla 12 ofrece una breve descripción de los atributos relacionados con la entidad Patrocinio. Se observan las características específicas y relevantes que definen la representación de los patrocinios dentro del prototipo.

*Tabla 13. Atributos para la entidad Producto*

<b>Atributo</b>	<b>Descripción</b>
Identifizer	Identificador del producto
Date	Fecha
Versión	Versión del producto
Authors	Autores del producto
Type	Tipo de producto
Language	Idioma del producto
Publisher	Editor del producto

La Tabla 13 proporciona una descripción concisa de los atributos asociados a la entidad Producto. En este contexto, se detallan de manera breve y específica las características relevantes que definen la representación de productos dentro del prototipo o sistema considerado.

La adopción de metadatos de Schema.org y OpenAIRE facilitan la representación de las entidades CERIF, lo cual como se ha mencionado contribuye a la interoperabilidad y el intercambio de información entre diferentes sistemas y plataformas, promoviendo así la integración fluida de datos de investigación. Además, aportan en la consistencia y la comprensión común de los datos, lo que resulta fundamental para la colaboración y la interoperabilidad en el panorama de la investigación científica y académica.

Con base en las consideraciones anteriores, se ha propuesto el modelo lógico que se presenta en la Figura 9 el cual permite la integración entre DSpace y CRIS. Considerando los requisitos específicos del contexto institucional de la Universidad del Cauca, en especial en el Grupo de Investigación en Inteligencia Computacional (GICO), mediante el que se logra una integración sinérgica que potencia la gestión de datos de investigación con los procesos, elementos y actores de la investigación. Este enfoque se traduce en un prototipo adaptado a las necesidades particulares, ofreciendo una solución para la gestión integrada de repositorios y sistemas CRIS en el ámbito académico.

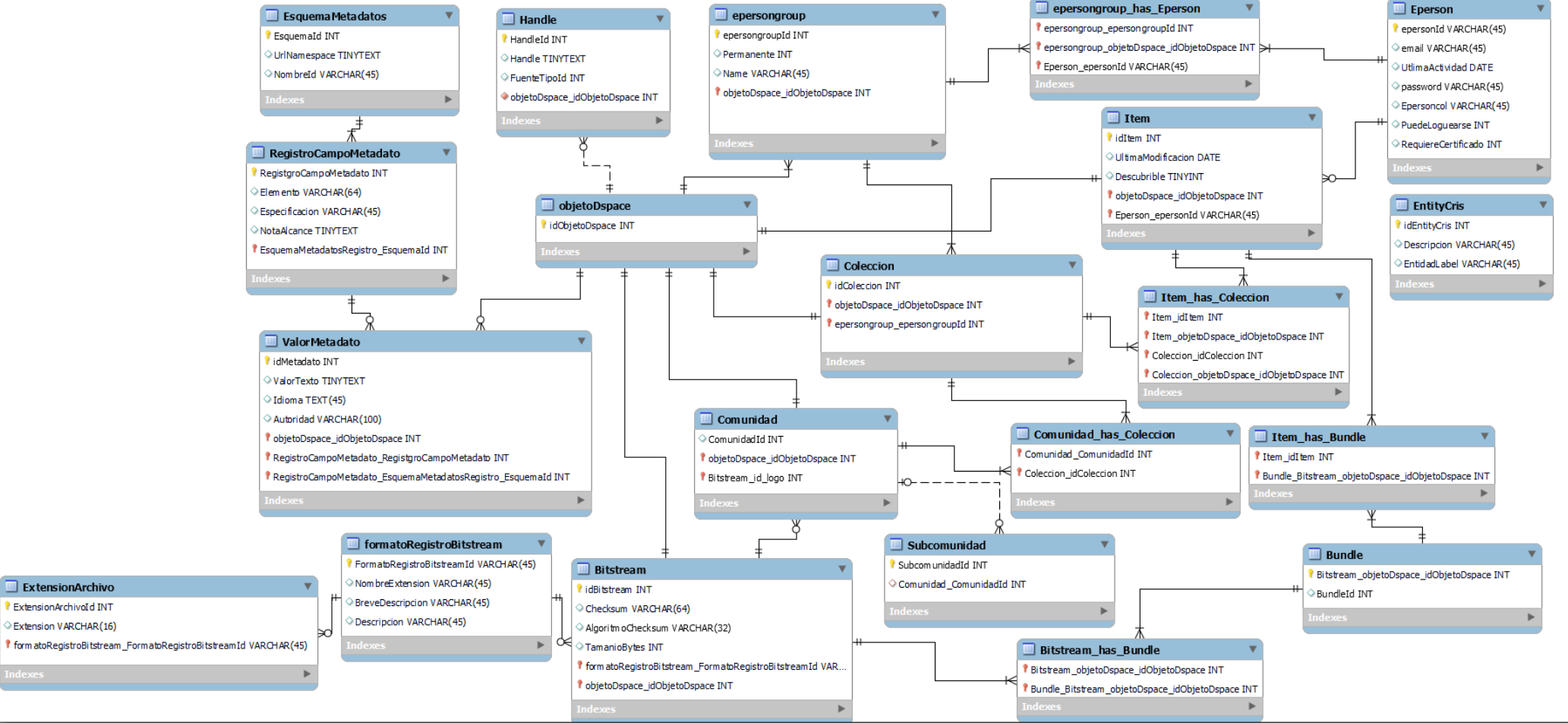


Figura 9. Modelo lógico DSpace-CRIS para INVESTCAUCA

En la Figura 9 se observa el modelo lógico de bases de datos, ofreciendo una representación visual de la estructura y las relaciones entre los distintos elementos de la base de datos.

Tras la definición de modelos y características apropiadas para la integración de sistema de gestión de datos de investigación, el siguiente paso de este trabajo se materializa en la presentación del siguiente capítulo. Donde, se trasladan los conceptos teóricos y las directrices establecidas en los modelos previos a la práctica mediante la implementación de un prototipo funcional. El objetivo es evidenciar cómo las teorías y los estándares debidamente definidos se traducen en un sistema concreto que atiende las necesidades específicas del grupo de investigación GICO de la Universidad del Cauca

## 4. PROTOTIPO DE REPOSITORIO DE DATOS DE INVESTIGACIÓN PARA LA UNIVERSIDAD DEL CAUCA INVESTUNICAUCA

El prototipo que se presenta a continuación se establece con un enfoque meticuloso y alineado con los estándares definidos, además, de los principios discutidos, incorporando elementos esenciales como metadatos, entidades y relaciones que han sido seleccionados y configurados.

### 4.1 Marco de trabajo para el desarrollo del prototipo

Para implementar el prototipo, se utilizó como metodología de desarrollo de software, Scrum, un marco ágil reconocido en el desarrollo de software que promueve la flexibilidad, colaboración y entregas incrementales. A continuación, se detallan los roles, eventos y artefactos específicos utilizados en este contexto:

#### 4.1.1 *Backlog* del Producto

##### 4.1.1.1 Instalación del *Backend*:

- Configuración del servidor y entorno.
- Instalación de la base de datos.
- Configuración de servicios de backend.

##### 4.1.1.2 Instalación del Frontend:

- Configuración de servidores web o de aplicaciones.
- Configuración de rutas y enlaces.
- Configuración de las dependencias del frontend.

##### 4.1.1.3 Adaptación de Metadatos:

- Definición de los campos de metadatos requeridos.
- Configuración de los esquemas de metadatos.

##### 4.1.1.4 Configuración de entidades CRIS:

- Definición de las entidades CRIS requeridas.
- Configuración de entidades CRIS.

##### 4.1.1.5 Adaptación de Formatos de Recursos:

- Configuración de formatos de archivos admitidos.

#### 4.1.1.6 Adaptación de Licencias:

- Configuración de opciones de licencia.

#### 4.1.1.7 Personalización de Interfaz de Usuario:

- Diseño y desarrollo de la interfaz de usuario.
- Personalización de la apariencia y la experiencia del usuario.

#### 4.1.2 Definición de Sprint

Se define un conjunto de cuatro (4) Sprints, los cuales tendrán una duración mayor a cuatro (4) semanas, esto ante la adaptación del marco de trabajo al contexto del proyecto. Este enfoque de planificación y ejecución se diseña cuidadosamente para optimizar la eficiencia y la productividad. Es importante destacar que el equipo encargado de llevar a cabo este proyecto consta de dos (2) personas, ambas estudiantes. La combinación de la definición clara de los Sprints y la gestión por parte de un equipo de dos miembros constituye los fundamentos sobre los cuales se desarrollará y ejecutará este proyecto de manera efectiva.

Tabla 14. Planificación de Sprints y Tareas

<b>Sprint</b>	<b>Product Backlog</b>
Sprint 1: Configuración Inicial	<ul style="list-style-type: none"><li>• Instalación del <i>Backend</i>.</li><li>• Instalación del <i>Frontend</i>.</li></ul>
Sprint 2: Fundamentos del Repositorio	<ul style="list-style-type: none"><li>• Adaptación de Metadatos.</li><li>• Adaptación de CRIS (Configuración inicial de entidades CRIS).</li></ul>
Sprint 3: Preparación de Datos	<ul style="list-style-type: none"><li>• Adaptación de Formatos de Recursos.</li><li>• Adaptación de Licencias.</li></ul>
Sprint 4: Interconexión y Personalización	<ul style="list-style-type: none"><li>• Personalización de Interfaz de Usuario (Diseño inicial y estructura).</li></ul>

En la Tabla 14, cada fila representa un sprint específico, enumerado desde el Sprint 1 hasta el Sprint 4, y detalla las tareas clave que se abordarán en cada iteración. En el Sprint 1, se destacan actividades relacionadas con la configuración inicial, como la instalación del *backend* y *frontend*. El Sprint 2 se centra en los fundamentos del repositorio, abordando la adaptación de metadatos y la configuración inicial de entidades CRIS. En el Sprint 3, la atención se dirige a la preparación de datos, con tareas que incluyen la adaptación de formatos de recursos y licencias. Finalmente, el Sprint 4 aborda aspectos de interconexión y personalización, como la personalización de la interfaz de usuario



## **4.2 Definición del prototipo de repositorio de datos de investigación**

### **INVESTUNICAUCA**

El repositorio de datos de investigación INVESTUNICAUCA, se define como un proyecto piloto para preservar los datos de investigación generados durante los diferentes procesos investigativos desarrollados en el Grupo de Investigación en Inteligencia Computacional (GICO) de la Universidad del Cauca, dado que este proyecto pertenece a la línea de investigación: Información y Tecnología de este grupo, vale destacar que este prototipo se define a partir de características que se consideran relevantes y señaladas en la revisión sistemática, las cuales se describen a continuación.

Para la definición legal del uso de los recursos digitales del prototipo INVESTUNICAUCA, se seleccionan herramientas para el licenciamiento y derechos de autor, esta elección se hace teniendo en cuenta las herramientas más utilizadas y mencionadas en la revisión sistemática de uso gratuito, donde se destacan las licencias *Creative Commons*, y la licencia Dominio público, esto también se define teniendo en cuenta las recomendaciones encontradas en la revisión sistemática en relación a lograr tener acceso a los recursos con la menor dificultad posible teniendo en cuenta la atribución al autor.

Las licencias *Creative Commons* en particular se definen como las licencias de uso recomendado para el prototipo al ser un conjunto de herramientas con características particulares que pueden satisfacer diferentes necesidades, entre las licencias CC se resaltan las licencias CC BY, la cual “permite que otros distribuyan, adapten y desarrollen su trabajo, incluso comercialmente, siempre que se atribuya el debido crédito por la creación original. Esta es la menos restrictiva de las licencias que se ofrecen. Recomendado para máxima difusión y uso de materiales licenciados.” (Commons, n.d.). Igualmente, la licencia CC BY-ND la cual “permite que otros reutilicen el trabajo para cualquier fin, incluido el comercial; sin embargo, no se puede compartir con otros en forma adaptada y se le debe proporcionar crédito.” (Commons, n.d.)

### **4.3 Perfil de aplicación del prototipo INVESTUNICAUCA**

Se decide utilizar un perfil de aplicación considerando que las necesidades del prototipo pueden satisfacerse de manera efectiva mediante la utilización de múltiples estándares de metadatos, en este sentido se analizan las utilidades de los estándares más mencionados en la revisión sistemática que son Dublin Core, DataCite y DDI, de estos estándares se decide utilizar Dublin Core por su utilidad para describir recursos digitales y DataCite por su utilidad para datos científicos y de investigación.

El perfil de aplicación para el prototipo INVESTUNICAUCA se presenta en la siguiente Tabla:

*Tabla 15. Perfil de aplicación para el prototipo INVESTUNICAUCA*

<b>Metadatos</b>	<b>Término utilizado</b>	<b>Descripción</b>
título	dc.title	El nombre dado a un recurso.
creador	dc.creator	Una entidad principalmente responsable de hacer el recurso.
sujeto	dc.subject	Un tema del recurso.
descripción	dc.description	La descripción del contenido del recurso.
editor	dc.publisher	Una entidad responsable de hacer que el recurso esté disponible.
colaborador	dc.contributor	Una entidad responsable de hacer contribuciones al recurso.
fecha	dc.date	Un punto o período de tiempo asociado con un evento en el ciclo de vida del recurso.
tipo	dc.type	La naturaleza o género del recurso.
formato	dc.format	El formato de archivo, el medio físico o las dimensiones del recurso.
identificador	dc.identifier	Una referencia inequívoca al recurso dentro de un contexto dado.
cobertura	dc.coverage	La extensión o ámbito del contenido del recurso.
licencia	dc.license	Un documento legal que da permiso oficial para hacer algo con el recurso.
derechos	dc.derechos	Información sobre los derechos que se tienen en y sobre el recurso.
identificador	DataCite.identifier	Un identificador único y persistente para el recurso, generalmente en forma de un DOI (Digital Object Identifier).
tipoidentificador	DataCite.identifierType	Tipo de identificador único.
versión	DataCite.Version	La versión específica del recurso.

*Fuente: Elaborado por los autores*

En la Tabla 15 se muestra el perfil de aplicación para el prototipo INVESTUNICAUCA, detallando el nombre del metadato a utilizar, su término según el estándar y una descripción del metadato correspondiente.

#### **4.4 Aspectos tecnológicos**

Se explora la dimensión tecnológica de la propuesta INVESTUNICAUCA, es decir, se analizan las herramientas y componentes esenciales que respaldarán la implementación eficiente y sostenible del prototipo del repositorio de datos de investigación. Esta sección se enfoca en aspectos como infraestructura, plataformas y estándares tecnológicos, asegurando así la funcionalidad integral del repositorio.

#### 4.4.1 Software

Para la propuesta del prototipo y con el fin de implementar un repositorio de datos de investigación con elementos CRIS, se decide hacer uso de la herramienta DSpace, se opta por esta alternativa en lugar de Invenio, Dataverse y CKAN, con base a estudios como el presentado en (Rodrigues Rufino Campêlo & Barreto Neto, 2019), donde se hace una comparación de los softwares libres previamente mencionados (Invenio, DSpace, CKAN y Dataverse), en ese estudio se identificaron ventajas que ofrece DSpace para la construcción de un repositorio de datos de investigación. Entre las ventajas se encuentra un mayor soporte de formatos admitidos, así como la adopción del estándar de metadatos Dublín Core, documentación detallada y una amplia comunidad de usuarios, lo que facilita su uso e implementación. Además, de que es de código abierto, enfocado a repositorios de datos de investigación y con amplia adopción en instituciones académicas y científicas.

De igual manera, DSpace se destaca por ser una plataforma robusta en lo que respecta a almacenar grandes volúmenes de datos y capacidad integral de gestión de datos, confiable, altamente adaptable lo cual la convierte en una plataforma con las facilidades para construir funcionalidades que son significativas para el usuario final, es una plataforma que permite una continua mejora y disponibilidad de soporte técnico con una comunidad de usuarios y desarrolladores activa (Verma & Kumar, 2018).

Cabe destacar que hasta la fecha (20/07/2023), DSpace en sus versiones más recientes cuenta con las versiones 7.X, en las cuales se incluye tecnologías actuales y avanzadas como lo es Angular, así como una nueva y completa REST API, utilizando las mejores prácticas modernas. Por ende, estas características hacen que DSpace se posicione como una plataforma de vanguardia y sea perfectamente adaptable de acuerdo con las necesidades del prototipo como a las particularidades de los datos de investigación.

La selección de la herramienta DSpace para implementación del prototipo también se hace teniendo en cuenta la extensión gratuita de código abierto de DSpace denominada DSpace-CRIS que según lo presentado en la página oficial de esta extensión “La principal característica de DSpace-CRIS es su modelo de datos flexible, que le permite recopilar y administrar datos e información de investigación típicos de un sistema CRIS, para definir entidades y atributos con sus enlaces recíprocos.” (DSpace-CRIS, 2023).

Es de resaltar que en la revisión sistemática se destaca que el CRIS comercial de mayor difusión internacional es PURE y que el CRIS no comercial de mayor difusión internacional es DSpace-CRIS, por esta razón también se define la utilización de la herramienta DSpace en su extensión DSpace-CRIS para la implementación del prototipo INVESTUNICAUCA, al ser una herramienta ampliamente utilizada en el

mundo, gratuita y que permite almacenar y administrar datos sobre la investigación realizada en una institución, puntualmente datos e información de investigación típicos de un sistema CRIS.

#### 4.4.2 Definición de formatos de archivos el prototipo INVESTUNICAUCA

Según el contexto del grupo de investigación GICO, se hace un análisis de los formatos o extensiones de los recursos o archivos que pueden ser utilizados en las tres líneas de investigación que manejan actualmente para tenerlos en cuenta en los formatos definidos en el prototipo INVESTUNICAUCA.

Dicha selección de formatos debe centrarse en las necesidades específicas de la investigación y en la disciplina correspondiente. No obstante, es importante asegurar que los datos sean interoperables y susceptibles de ser reutilizados de manera sencilla. Por lo tanto, se aconseja la utilización de formatos de archivo abiertos en lugar de los cerrados como una práctica recomendada. Se sugiere otorgar preferencia a aquellos que cuenten con un amplio reconocimiento en la comunidad de investigación. Por ejemplo, en el caso de los datos tabulares, es más adecuado compartirlos en formato CSV en lugar de utilizar archivos XLS (Excel).

Hay que destacar que los datos de investigación pueden compartirse en una variedad de formatos y tipos de contenido, algunos de los cuales se mencionan a continuación en la Tabla 16.

*Tabla 16. Formatos que soporta la plataforma software y que son requeridos para prototipo INVESTUNICAUCA.*

<b>Categoría</b>	<b>Formatos</b>
<b>Multimedia</b>	jpeg, png, gif mp3, wav, mp4, avi, mov
<b>Texto</b>	doc, docx, odt, txt, rtf, pdf, html, xml, ppt, pptx
<b>Software</b>	java, python, c++, r, exe, jar, python
<b>Datos y Tablas</b>	csv, xls, xlsx, sql, sqlite
<b>Archivos Comprimidos</b>	zip, tar, gz, 7z

La Tabla 16 proporciona información estructurada sobre los diferentes formatos o tipos de archivos que soporta la plataforma software y que son requeridos para el prototipo INVESTUNICAUCA.

## 4.5 Implementación del Prototipo

A través del proceso de implementación, se busca evaluar el modelo propuesto en un entorno real, lo que permitirá validar la aplicabilidad de los conceptos teóricos analizados, para comprender cómo los estándares, licencias y características pueden operar de manera práctica en la gestión de datos de investigación.

### 4.5.1 Precondiciones.

Para la instalación y configuración del prototipo se deben tener en consideración una serie de precondiciones tanto para la instalación del *backend* como del *frontend* del aplicativo, las cuales se enumeran a continuación:

Tabla 17. Herramientas software para realizar la instanciación de DSpace-CRIS

Software	Versión	Descripción
JDK kit 11	11.0.18	Proporciona el entorno necesario para el desarrollo y ejecución de aplicaciones Java
Apache mvn	3.6.3	Utilizado para la gestión de proyectos Java, Maven simplifica la compilación y manejo de dependencias.
Ant	1.10.13	Similar a Maven que se utiliza para automatizar procesos de construcción de software.
PostgreSQL	13.8	Sistema de gestión de bases de datos relacional de código abierto.
Apache Solr	8.11.2	Se utiliza para indexar, buscar y analizar eficientemente grandes cantidades de datos textuales.
Apache Tomcat	9	Servidor web y contenedor <i>servlet</i> de código abierto. La versión 9 se implementa para alojar y ejecutar aplicaciones web Java.
Node.js	v18.14.1	Se utiliza en el <i>frontend</i> para ejecutar código JavaScript.
Npm	9.3.1.	Administrador de paquetes oficial de Node.js, facilitando la gestión de dependencias.
Git	2.41.0	Para colaborar en el desarrollo y mantener un registro ordenado de las modificaciones realizadas.

La Tabla 17 se detalla las herramientas de software utilizadas para llevar a cabo la instanciación o instalación de DSpace-CRIS. Incluye el nombre del recurso de software y la versión correspondiente.

### 4.5.2 Instalación

El código inicial para el prototipo, tanto para el *backend* como el *frontend* se toma de la página de DSpace-CRIS: (DSpace-CRIS, 2023)

El código base, se almacena en un repositorio público para poder ajustarlo según las necesidades y diseño del proyecto y así llevar un control de versiones de los ajustes del prototipo, el código del *backend* del prototipo se toma queda disponible en el siguiente repositorio de manera pública: <https://github.com/adrianbn96/Dspacebackunicauca> y el código del *fronted* del prototipo queda disponible en el siguiente repositorio de manera pública: <https://github.com/adrianbn96/Dspacefrontunicauca>

Es de resaltar que para la instalación de DSpace-CRIS se decide utilizar el sistema operativo Windows 10.

### 4.5.3 Persistencia

1. Para la instalación de PostgreSQL se define:

```
User = postgres  
Password = *****
```

2. En el archivo `pg_hba.conf` se agrega host DSpace

```
host dspace dspace 127.0.0.1 255.255.255.255 md5
```

Esta línea permite que el usuario "DSpace" pueda conectarse a la base de datos "DSpace" desde localhost (127.0.0.1) utilizando autenticación mediante contraseña MD5

3. Se debe crear el usuario DSpace y crear la base de datos DSpace y configurar super usuario para usuario de DSpace. Por medio de `psql Shell` se realizan las siguientes configuraciones:

- Crear usuario DSpace con contraseña DSpace y base de datos DSpace:

```
create user DSpace with password 'DSpace'  
create database DSpace encoding 'UNICODE';
```

- Establecer usuario DSpace como superuser

```
grant all privileges on database DSpace to DSpace;
```

- Crear extensión `pgcrypto` cho base de datos DSpace

```
create extension pgcrypto;
```

#### 4.5.4 Instalación *Backend*

Para la instalación del *backend* de DSpace-CRIS, se debe primero tener en cuenta configurar Apache tomcat, se define `username: DSpace` y `pass: DSpace`, también se debe configurar el archivo `server.xml`

1. Se deben buscar las siguientes líneas:

```
connectionTimeout="20000"  
redirectPort="8443" />
```

2. Se deben reemplazar por:

```
minSpareThreads="25"  
enableLookups="false"  
redirectPort="8443"  
connectionTimeout="20000"  
disableUploadTimeout="true"  
URIEncoding="UTF-8"/>
```

3. Se debe crear la variable `TOMCAT_USER` con el valor `dspace`
4. Una vez configurado Tomcat, se inicia la instalación con la creación de la carpeta `dspace` en el disco local `C:\dspace`
5. Se debe tener el código fuente del *backend* de DSpace-CRIS en el disco local `C:\`
6. Configuración del archivo `local.cfg`

En `C:\DSpace-DSpace-7.5\DSpace\config` se encuentra el archivo con nombre `local.cfg.EXAMPLE`, el cual se puede copiar y renombrar como `local.cfg`, este archivo se utiliza para definir y modificar las configuraciones específicas de una instalación de DSpace, en este archivo se realizan las siguientes configuraciones:

- Ruta instalación DSpace:

```
DSpace.dir=C:/DSpace
```

- Nombre de instalación DSpace

```
DSpace.name = DSpaceCris Unicauca
```

- Eliminar comentario en la línea de ruta **Solr**

```
solr.server = http://localhost:8983/solr
```

- Configuraciones persistencia

```
db.username = DSpace  
db.password = DSpace
```

7. Luego de tener las configuraciones en el archivo `local.cfg`, en la ruta `C:\DSpace-DSpace-7.5` se ejecuta en consola el comando `mvn package` para compilar y empaquetar el código fuente y los recursos del proyecto, al finalizar, en la ruta `C:\DSpace-DSpace-7.5\DSpace\target\DSpace-installer` ejecutar el comando `ant fresh_install` para la instalación completa y nueva del proyecto.
8. En la ruta `C:\DSpace\bin\`, ejecutar el comando `DSpace database migrate` para aplicar migraciones de base de datos.
9. Copiar el contenido de la ruta `C:\DSpace\webapps` en la ruta `C:\Program Files\Apache Software Foundation\Tomcat 9.0\webapps`
10. Copiar el contenido de la ruta `C:\DSpace\solr` en la ruta `C:\solr-8.11.2\server\solr\configsets`
11. Se debe ejecutar Solr y Tomcat, para iniciar Solr ejecutar el comando `solr start`.
12. Para crear cuenta de administrador de DSpace, en la ruta `C:\DSpace/bin/DSpace` se ejecuta el comando `create-administrator` y se definen las credenciales con las cuales posteriormente se va a ingresar como administrador al sistema.

A continuación, se presenta captura de pantalla de Tomcat y Solr en funcionamiento, donde se evidencia las peticiones *backend* ya disponibles después del proceso de instalación de DSpace-CRIS.



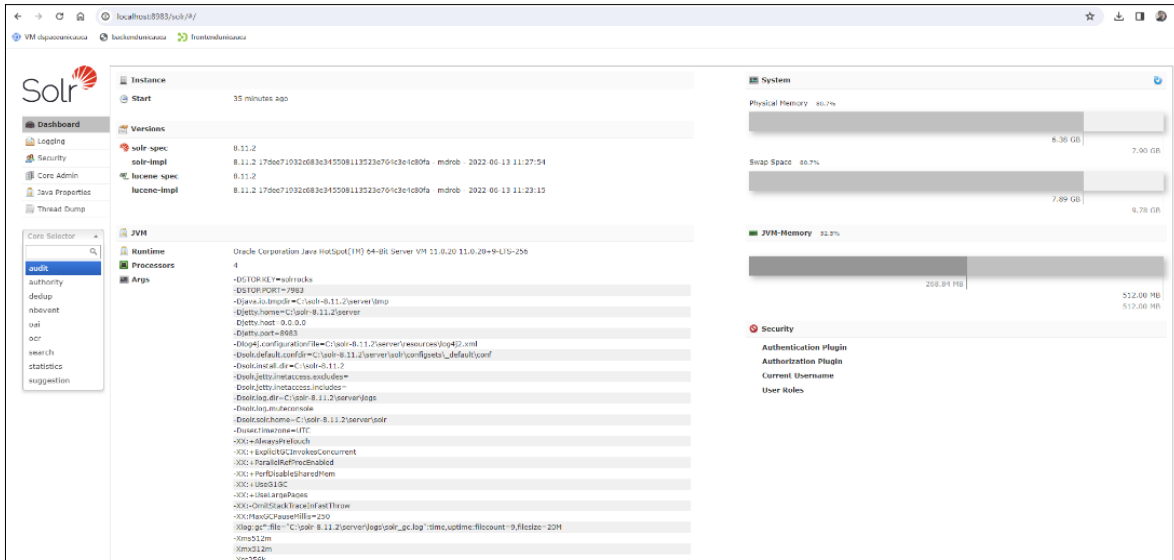


Figura 10. Interfaz gráfica de la plataforma de búsqueda Solr para el prototipo INVESTUNICAUCA.

En la Figura 10 se presenta el sistema Solr en funcionamiento, tras la realización de las configuraciones necesarias. Su despliegue operativo, reflejado en la imagen, destaca la implementación de las configuraciones necesarias para potenciar la búsqueda y análisis de grandes conjuntos de datos textuales, proporcionando así una herramienta clave para manejar la información académica e investigación eficiente.

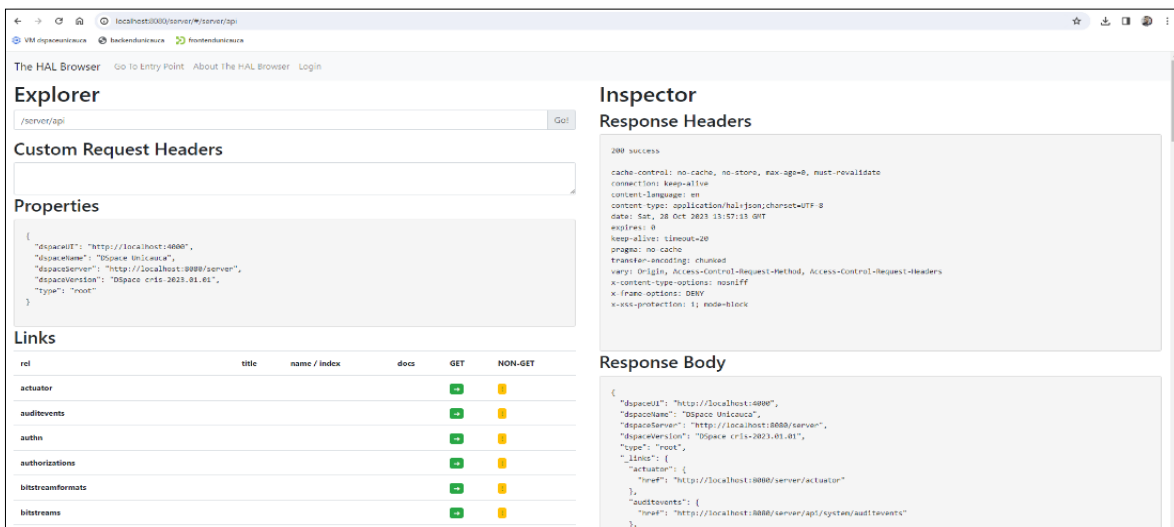


Figura 11. Interfaz gráfica del servidor web Tomcat para el prototipo INVESTUNICAUCA.

En la Figura 11 se puede observar el funcionamiento exitoso del servidor web Apache Tomcat, después de haber aplicado las configuraciones que se requiere. La representación gráfica refleja el éxito en la implementación de las configuraciones, asegurando así un despliegue operativo óptimo.

#### 4.5.5 Instalación *Frontend*

1. Teniendo en cuenta que ya se tiene instalado Node.js y npm, se debe instalar por medio de npm Yarn como administrador de paquetes para JavaScript y PM2 es una herramienta de gestión de procesos para aplicaciones Node.js con los siguientes comandos:

```
npm install --global yarn  
npm install --global pm2.
```

2. En la ruta del código fuente de DSpace angular se ejecuta para instalar las dependencias necesarias del proyecto el siguiente comando:

```
yarn install
```

3. En la misma ruta se ejecuta la siguiente instrucción para llevar a cabo la construcción o compilación de la aplicación.

```
yarn build:prod
```

4. En la ruta /DSpace-angular/config, existe un archivo ejemplo de configuración config.example.yml, el cual sirve como guía de configuración de la aplicación, para su uso en un entorno de producción se crea el archivo config.prod.yml tomando como base el archivo de ejemplo, en este archivo se define protocolo, puertos y host del api REST y la Interfaz de Usuario.
5. En la ruta raíz de DSpace angular se crea un archivo con nombre *DSpace-angular.json*, con el cual pm2 utiliza para administrar aplicaciones Node.js y se define lo siguiente.

```
{"apps": [{"name": "DSpace-angular", "cwd":  
"rutaDSpaceangular", "script":  
"dist\\server\\main.js", "instances": "max", "exec_mode":  
"cluster", "env": {"NODE_ENV": "production"}}]}
```

6. Para iniciar la aplicación angular con pm2, se ejecuta el siguiente comando:

```
pm2 start DSpace-angular.json
```

A continuación, en la Figura 12 se presenta captura de pantalla del *frontend* del prototipo en angular, donde se evidencia el *home page* de DSpace-CRIS, al cual se puede ingresar con las credenciales definidas en la instalación del *backend* para el rol administrador.

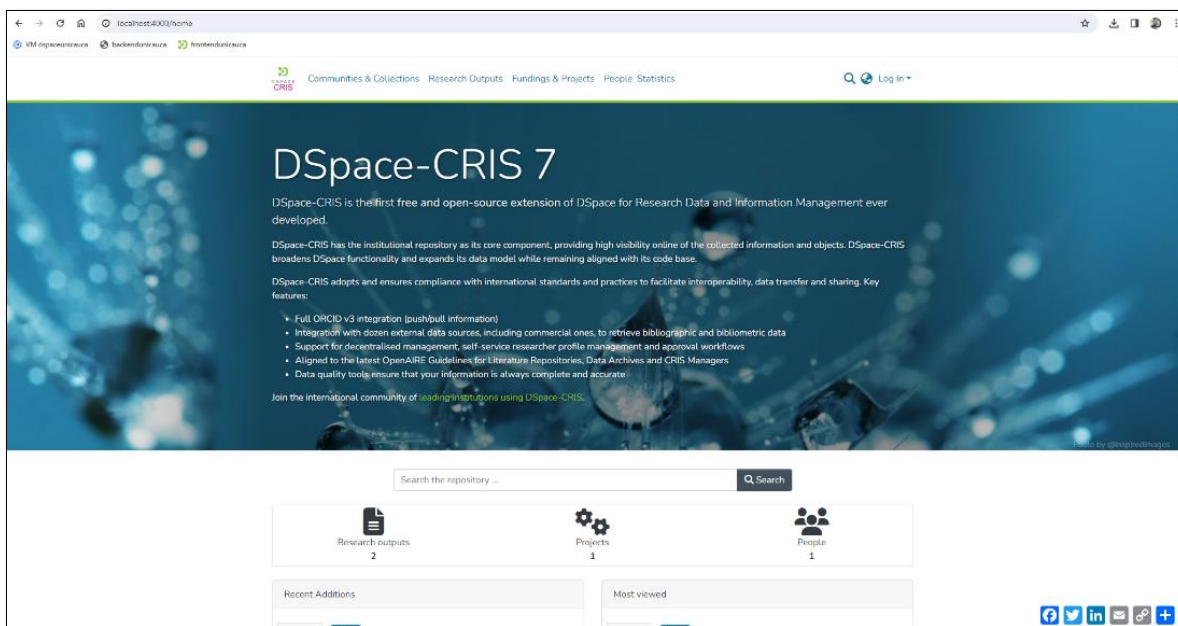


Figura 12. Captura de pantalla de la interfaz gráfica de la instalación *dspace-cris* para prototipo *INVESTUNICAUCA*.

En la Figura 12 se visualiza la interfaz activa, ofreciendo la posibilidad de consultar los recursos depositados. Es fundamental señalar que el acceso y la manipulación de recursos están condicionados por los permisos asignados, asegurando así la integridad y seguridad del sistema.

#### 4.6 Configuración del repositorio para un grupo de investigación.

Teniendo en cuenta que el prototipo se debe ajustar a las necesidades que se presentan dentro del Grupo de Investigación en Inteligencia Computacional (GICO), el cual pertenece al departamento de sistemas de la Universidad del Cauca, en la página web (*GICO | Grupo de Investigación En Inteligencia Computacional*, n.d.), se presentan algunas características relevantes sobre el grupo de investigación, entre las que se destacan que está reconocido por Colciencias en categoría B y que actualmente trabaja en 3 líneas de investigación que son Inteligencia Artificial y Educación, Ciencia de la Información y la Ciencia de la Computación y Ciencia de los datos.

En la adaptación del prototipo se tienen en cuenta las características principales encontradas en la revisión sistemática y considerando las necesidades que se

presentan en los diferentes procesos investigativos desarrollados en el Grupo de Investigación en Inteligencia Computacional (GICO) de la Universidad del Cauca, a continuación, se presentan las adaptaciones al prototipo.

Para la construcción y adaptación del prototipo INVESTUNICAUCA se siguió la metodología scrum para gestionar este proyecto. Los roles según la metodología Scrum son adaptados teniendo en cuenta que deben ser ejecutados por los integrantes del proyecto, tomando el rol *Product Owner* de manera compartida entre los dos estudiantes integrantes de este trabajo de grado (Guido Imbachi y Adrián Felipe Burbano), quienes definen las necesidades del prototipo según las recomendaciones de la directora del proyecto (Sandra Milena Roa Martínez), recomendaciones que también de manera compartida se desarrollan en el rol de equipo de desarrollo por los integrantes del proyecto, el rol de *Scrum Master* por parte de Adrián Felipe Burbano, quien lidera la apropiación de las prácticas Scrum.

#### **4.6.1 Instanciación del perfil de aplicación en el prototipo**

##### **INVESTUNICAUCA.**

La instalación de DSpace permite gestionar desde la interfaz de usuario los perfiles de metadatos que son el conjunto de campos de metadatos que siguen estándares específicos, como Dublin Core, Datacite, Schema.org, entre otros.

Para acceder a la opción de gestión de metadatos se debe ingresar al sistema con el rol de administrador. Después de haber realizado el inicio de sesión, en la barra lateral se encuentra la opción “Registro”, en donde se encuentra la opción de “Metadatos”, a continuación, se presenta captura de pantalla de la página de registro de metadatos en la instalación del prototipo.

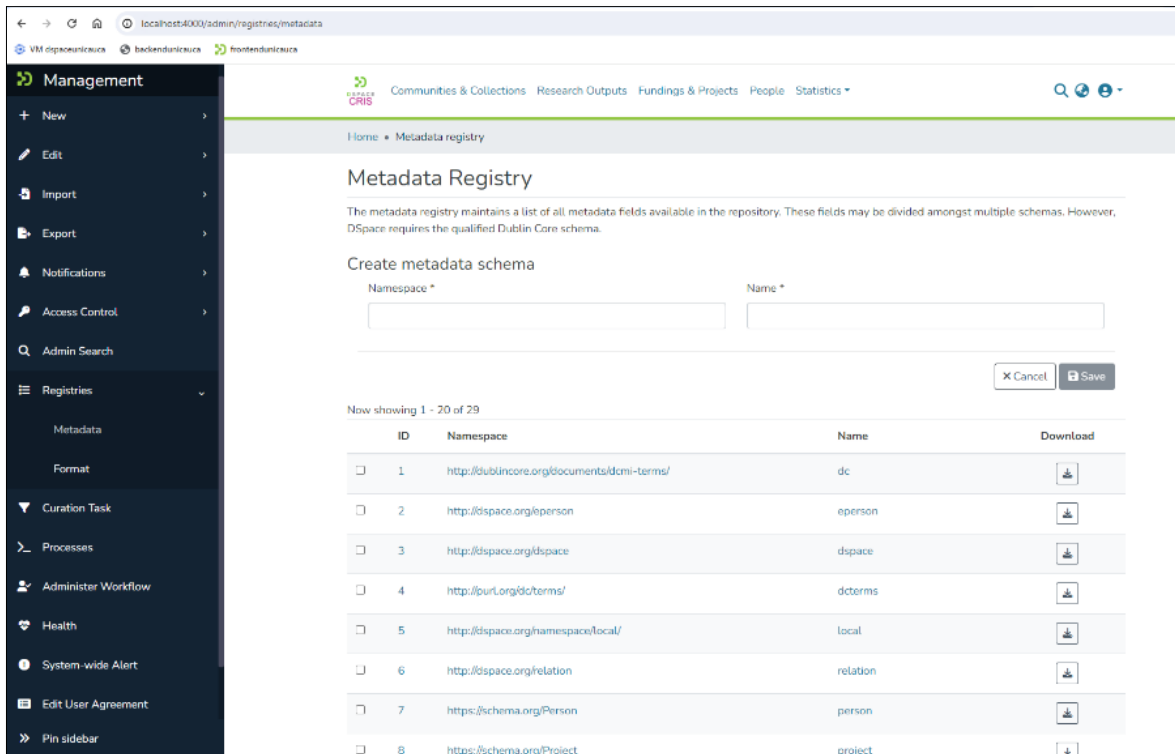
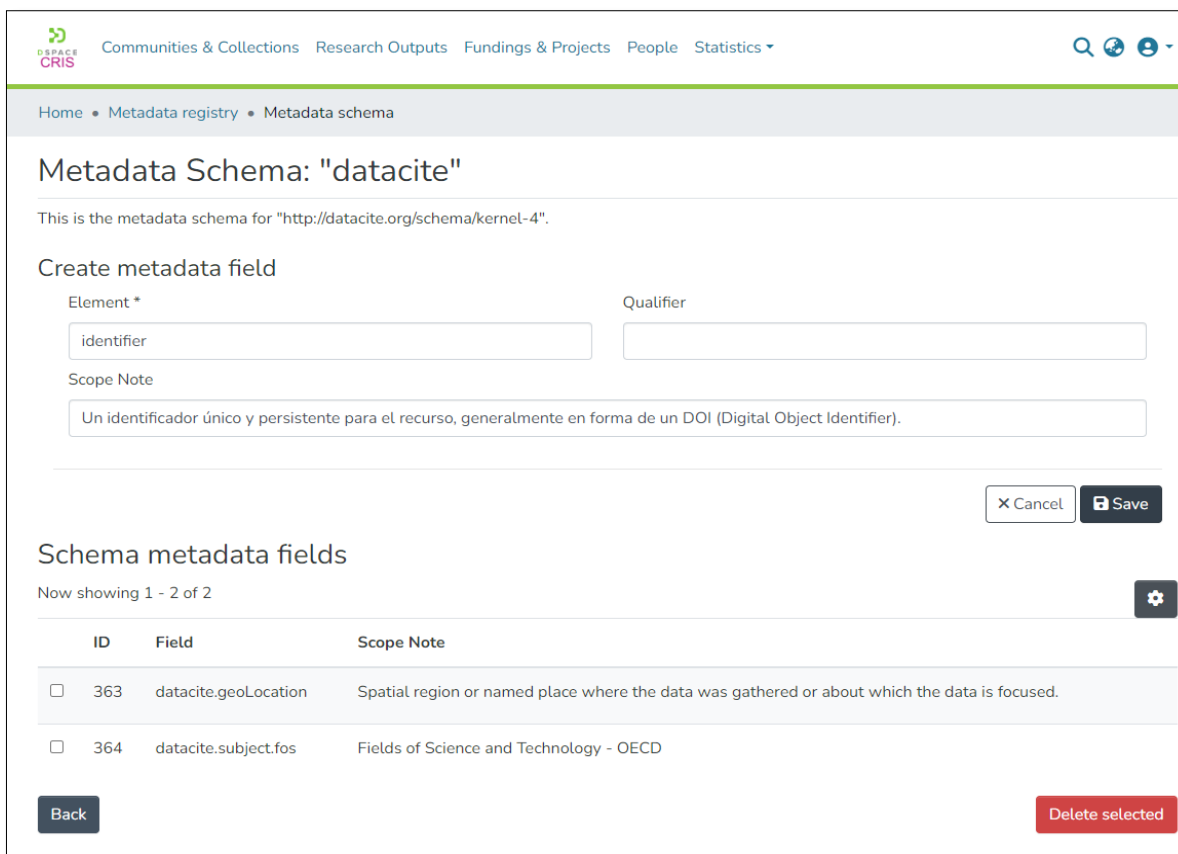


Figura 13. Captura de pantalla de la interfaz gráfica para definición de metadatos del prototipo INVESTUNICAUCA.

Cabe resaltar que por defecto en la página de gestión de metadatos del prototipo informa el mensaje “El registro de metadatos mantiene una lista de todos los campos de metadatos disponibles en el repositorio. Estos campos pueden estar divididos entre múltiples esquemas. Sin embargo, DSpace requiere el esquema calificado de Dublín Core.”, esto es relevante ya que Dublín Core es el estándar definido para el prototipo, y también se encuentra datacite por defecto, en estos esquemas de metadatos se encuentran una serie de campos correspondientes al esquema puntual, aunque ya viene con una serie de campos, también se pueden agregar un campo de metadatos faltante o si no se encuentra en los que trae por defecto.

A continuación, se muestra la creación de un campo nuevo en el esquema de metadatos Datacite.



The screenshot shows the 'Metadata Schema: "datacite"' page. At the top, there is a navigation bar with the Datacite logo and links for 'Communities & Collections', 'Research Outputs', 'Fundings & Projects', 'People', and 'Statistics'. Below the navigation bar, the breadcrumb trail reads 'Home • Metadata registry • Metadata schema'. The main heading is 'Metadata Schema: "datacite"'. A sub-heading states 'This is the metadata schema for "http://datacite.org/schema/kernel-4"'. The 'Create metadata field' section contains three input fields: 'Element \*' with the value 'identifier', 'Qualifier' (empty), and 'Scope Note' with the text 'Un identificador único y persistente para el recurso, generalmente en forma de un DOI (Digital Object Identifier)'. Below these fields are 'Cancel' and 'Save' buttons. The 'Schema metadata fields' section shows a table with two entries:

ID	Field	Scope Note
<input type="checkbox"/> 363	datacite.geoLocation	Spatial region or named place where the data was gathered or about which the data is focused.
<input type="checkbox"/> 364	datacite.subject.fos	Fields of Science and Technology - OECD

At the bottom of the table, there are 'Back' and 'Delete selected' buttons.

Figura 14. Captura de pantalla de la interfaz gráfica con la definición de metadatos datacite del prototipo investunicauca.

En la Figura 14, se observa la configuración del campo "*datacite.identifier*". Este campo no está incluido por defecto en la instalación del prototipo, sin embargo, está definido en el perfil de aplicación. De igual forma se agregan los campos faltantes que se definieron en el perfil de aplicación que son "*datacite.identifierType*" y "*datacite.version*".

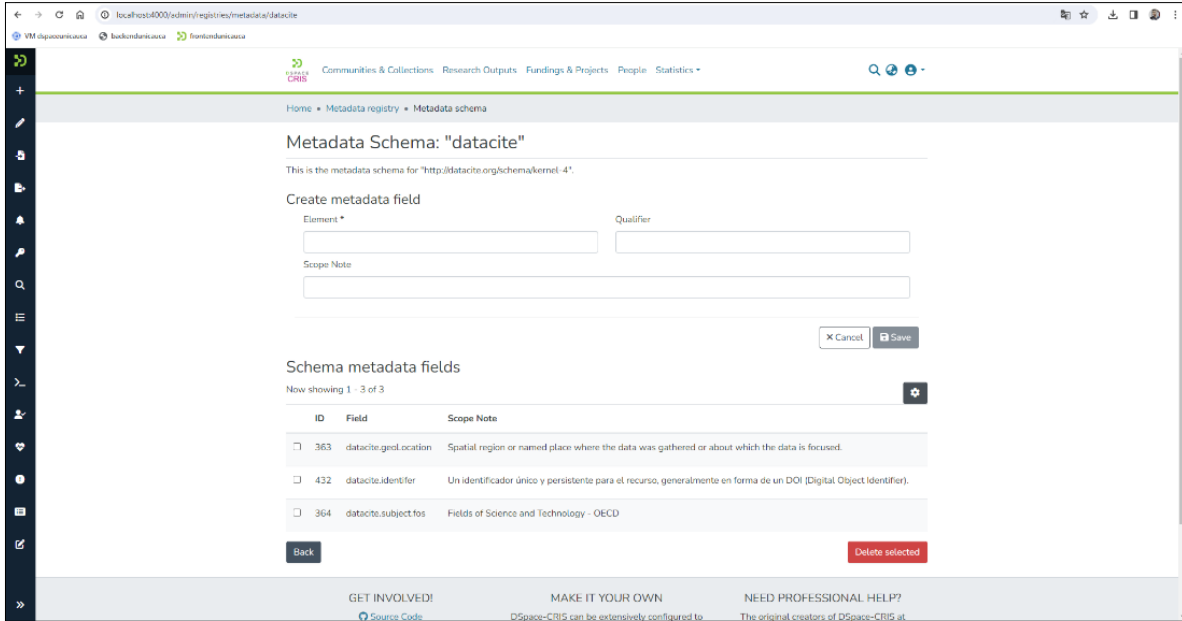


Figura 15. Captura de pantalla de la interfaz gráfica metadato agregado a datacite de prototipo investunicauca.

Como se muestra en la Figura 15, se define *Element* como valor requerido, que para este caso sería *identifier*, también *Qualifier* dado el caso que el nuevo campo tenga un calificador y *Scope Note* donde se describe brevemente la utilidad del nuevo campo a crear, y se selecciona la opción *save*. Luego de definir los campos necesarios, se guardan los datos y quedan disponibles como nuevos metadatos dentro del esquema de metadatos en particular.

Al definir el metadato *datacite.identifier* que hace parte del estándar de metadatos datacite, se puede configurar las colecciones para que, al crear un ítem de la colección, tengan este nuevo metadato como se muestra a continuación.

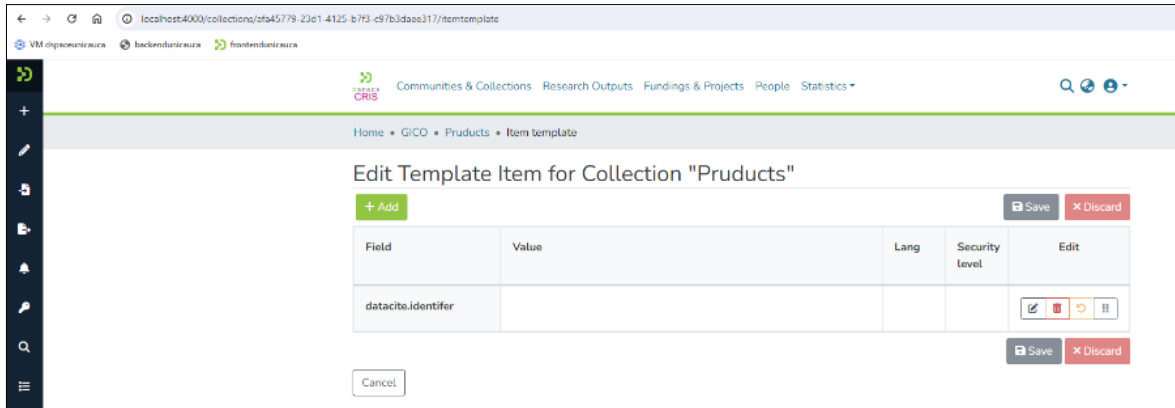


Figura 16. Captura de pantalla de la interfaz gráfica colección con nuevo metadato de prototipo INVESTUNICAUCA.

En la Figura 16 se evidencia la habilitación del nuevo metadato para su uso práctico y funcional en una variedad de ítems dentro de las respectivas colecciones. Esta incorporación asegura que el metadato esté disponible y sea fácilmente accesible, ofreciendo así una mayor flexibilidad y coherencia en la gestión de la información en el contexto de las diversas comunidades y colecciones pertinentes. Esta integración exitosa del nuevo metadato refleja un avance significativo en la mejora de la estructura y eficacia del sistema, facilitando la gestión integral de los datos asociados.

#### 4.6.2 Definición de identificadores persistentes para el prototipo del repositorio de datos de investigación INVESTUNICAUCA

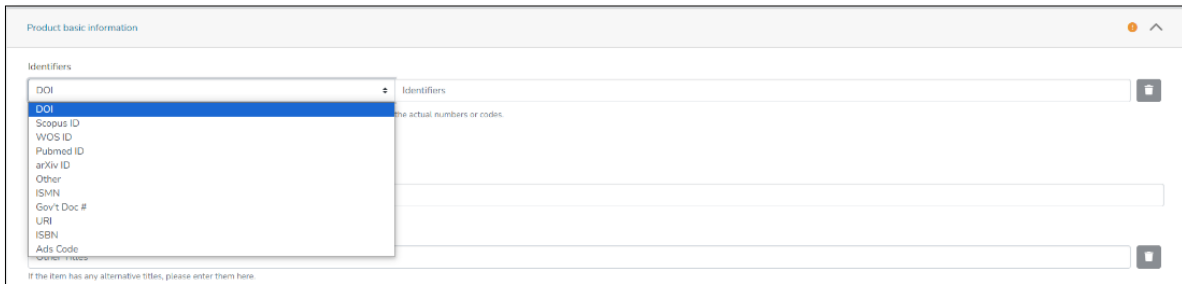
Los identificadores persistentes en el prototipo son definidos a partir de los metadatos de un estándar de metadatos, a continuación, se muestran los principales en el estándar Dublín Core, donde se encuentran gran variedad de identificadores, cabe resaltar que, al ser campos de metadatos, se pueden agregar más si hace falta alguno, en el caso de Dublín Core los tipos de identificadores persistentes son calificadores de “*dc.identifier*”, estos se ven reflejados en la creación de ítems en el desplegable con nombre “*Identifiers*”.



<input type="checkbox"/>	32	dc.identifier.doi	
<input type="checkbox"/>	26	dc.identifier.govdoc	A government document number
<input type="checkbox"/>	27	dc.identifier.isbn	International Standard Book Number
<input type="checkbox"/>	35	dc.identifier.isi	Web of Knowledge Identifier
<input type="checkbox"/>	30	dc.identifier.ismn	International Standard Music Number
<input type="checkbox"/>	28	dc.identifier.issn	International Standard Serial Number
<input type="checkbox"/>	31	dc.identifier.other	A known identifier type common to a local collection.

*Figura 17. Captura de pantalla de la interfaz gráfica metadatos de identificadores persistentes en la creación de un ítem del prototipo INVESTUNICAUCA.*

Los identificadores persistentes, visibles en la Figura 17, ahora podrán ser elegidos de manera conveniente a través de un menú desplegable en los formularios de creación de distintos ítems.



*Figura 18. Captura de pantalla de la interfaz gráfica opciones de identificadores persistentes en la creación de un ítem del prototipo INVESTUNICAUCA.*

En la Figura 18, se evidencia que el identificador deseado se podrá seleccionar en el menú desplegable, convirtiéndose así en un metadato asociado al ítem después de que este haya sido depositado o guardado.

### 4.6.3 Definición de entidades CERIF para el prototipo de repositorio de datos de investigación INVESTUNICAUCA

Para la definición de las entidades CRIS, que son las definidas en el modelo CERIF, se hace por medio de una utilidad del aplicativo que se ejecuta con el comando:

```
DSpace dsrun org.DSpace.app.util.InitializeEntityTypesOnly -d,
```

Las relaciones entre las identidades por medio de los archivos *correction-relationship-types.xml* y *hide-sort-relationship-types.xml* que son archivos de configuración utilizado para definir los tipos de relaciones entre los elementos en un repositorio digital basado en DSpace, más precisamente en DSpace-CRIS.

```

correction-relationship-types.xml X
dSPACE > config > entities > correction-relationship-types.xml
1  <?xml version="1.0" encoding="UTF-8"?>
2  <!DOCTYPE relationships SYSTEM "relationship-types.dtd">
3
4  <relationships>
5    <type>
6      <leftType>Person</leftType>
7      <rightType>Person</rightType>
8      <leftwardType>isCorrectionOfItem</leftwardType>
9      <rightwardType>isCorrectedByItem</rightwardType>
10     <leftCardinality>
11       <min>0</min>
12     </leftCardinality>
13     <rightCardinality>
14       <min>0</min>
15     </rightCardinality>
16   </type>
17   <type>
18     <leftType>Project</leftType>
19     <rightType>Project</rightType>
20     <leftwardType>isCorrectionOfItem</leftwardType>
21     <rightwardType>isCorrectedByItem</rightwardType>
22     <leftCardinality>
23       <min>0</min>
24     </leftCardinality>
25   </type>
hide-sort-relationship-types.xml X
dSPACE > config > entities > hide-sort-relationship-types.xml
1  <?xml version="1.0" encoding="UTF-8"?>
2  <!DOCTYPE relationships SYSTEM "relationship-types.dtd">
3
4  <relationships>
5    <type>
6      <rightType>Person</rightType>
7      <leftwardType>isResearchOutputsHiddenFor</leftwardType>
8      <rightwardType>notDisplayingResearchOutputs</rightwardType>
9      <leftCardinality>
10     <min>0</min>
11   </leftCardinality>
12   <rightCardinality>
13     <min>0</min>
14   </rightCardinality>
15 </type>
16 <type>
17   <rightType>Person</rightType>
18   <leftwardType>isProjectsHiddenFor</leftwardType>
19   <rightwardType>notDisplayingProjects</rightwardType>
20   <leftCardinality>
21     <min>0</min>
22   </leftCardinality>
23   <rightCardinality>
24     <min>0</min>

```

Figura 19. Captura de pantalla de archivos para definición de entidades CERIF en prototipo INVESTUNICAUCA.

En la Figura 19, se presenta una visualización de los archivos configurados, donde se definen las entidades CERIF. Esta representación no solo abarca la configuración de las entidades, sino que también destaca la definición y configuración de las relaciones entre estas diversas entidades dentro del prototipo propuesto.

#### 4.6.4 Definición y estandarización de formatos de archivos para el prototipo del repositorio de datos de investigación INVESTUNICAUCA

Para la definición de los formatos que serán considerados según el diseño del prototipo, en la opción de “registros”, en el apartado de “*Formats*”, se encuentra una serie de formatos predefinidos más utilizados, también se encuentra la opción de agregar nuevos formatos.

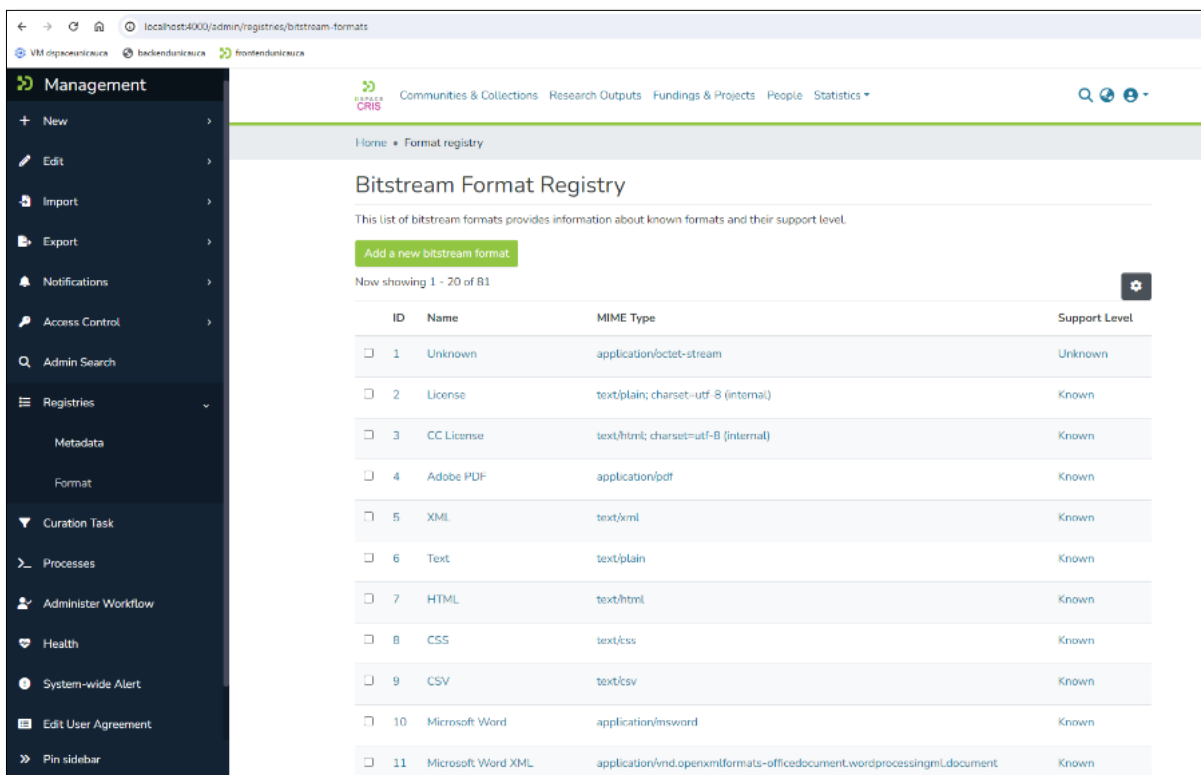


Figura 20. Captura de pantalla de formatos de archivos en prototipo INVESTUNICAUCA.

En la Figura 20, se observa que entre los formatos que se encuentran por defecto en la instalación inicial se encuentran formatos como PDF, XML, HTML, CSS, CSV, JPEG, GIF, RDF XML, MP3, entre otros.

## 4.6.5 Definición de licencias para el prototipo del repositorio de datos de investigación INVESTUNICAUCA

La definición de las licencias que según el mapeo sistemático serían las licencias *Creative commons*, se realizan en el archivo de configuración de DSpace, el cual es `DSpace.cfg`, donde existen una serie de configuraciones que permiten configurar y mostrar en la interfaz gráfica el paquete de licencias *creative commons*.

```
dspace.cfg x
dspace > config > dspace.cfg
1007 #-----#
1008 #-----UI-Related CONFIGURATIONS-----#
1009 #-----#
1010 # These configs are used primarily by the User Interface #
1011 #-----#
1012
1013 # Determine if super administrators (those whom are in the Administrators group)
1014 # can login as another user from the "edit eperson" page. This is useful for
1015 # debugging problems in a running dspace instance, especially in the workflow
1016 # process. The default value is false, i.e. no one may assume the login of another user.
1017 #webui.user.assumelogin = true
1018
1019 #allow admin to impersonate another admin user
1020 #webui.user.assumelogin.admin = true
1021
1022 # whether to display the contents of the licence bundle (often just the deposit
1023 # licence in standard DSpace installation)
1024 # TODO: UNSUPPORTED in DSpace 7.0
1025 webui.licence_bundle.show = true
1026
1027 ##### Hide Item Metadata Fields #####
1028 # Fields named here are hidden in the following places UNLESS the
1029 # logged-in user is an Administrator:
1030 # 1. REST API (and therefore UI)
1031 # 2. RDF (every where as there is currently no possibility to authenticate)
1032 # 3. OAI (every where as there is currently no possibility to authenticate)
1033 # Attention: You need to rebuild the OAI SOLR index after every change of
1034 # this property. Run [dspace-install]/bin/dspace oai import -c to do so.
1035 #
1036 # To designate a field as hidden, add a property here in the form:
1037 # metadata.hide.SCHEMA.ELEMENT.QUALIFIER = true
1038 #
1039 # This default configuration hides the dc.description.provenance field,
1040 # since that usually contains email addresses which ought to be kept
1041 # private and is mainly of interest to administrators:
1042 metadata.hide.dc.description.provenance = true
1043
1044 ##### Settings for Submission Process #####
1045
1046 #Default relationship type in submission
1047 submission.default.entitytype = Publication
1048
1049 # Whether or not we REQUIRE that a file be uploaded
1050 # during the 'Upload' step in the submission process
1051 # Defaults to true; If set to 'false', submitter has option to skip upload
1052 #webui.submit.upload.required = true
1053
1054 # Which field should be used for type-bind
1055 # Defaults to 'dc.type'; If changing this value, you must also update the related
1056 # dspace-angular environment configuration property submission.typeBind.field
1057 #submit.type-bind.field = dc.type
1058
```

Figura 21. Captura de pantalla de archivos para configuración de licencias *creative Commons* en prototipo *INVESTUNICAUCA*.

En la versión de DSpace 7.0 se informa que no existe soporte para trabajar con estas licencias, por esta razón se deja descrito que se debería definir la variable `webui.licence_bundle.show = true` para cuando sea soportado por versiones posteriores.

DSpace-CRIS viene con una licencia de depósito por defecto (licencia GNU *General Public License* (GPL) versión 3) la cual es descrita de forma clara, informando a las condiciones que se aceptan al confirmar esta licencia para el recurso a depositar, cabe resaltar que es requerida al momento de crear un ítem, teniendo en cuenta que también es requerido adjuntar un archivo al ítem que se desea crear y esta licencia es precisamente para definir y comunicar los términos y condiciones bajo los cuales se pueden utilizar y compartir el depósito a realizar.

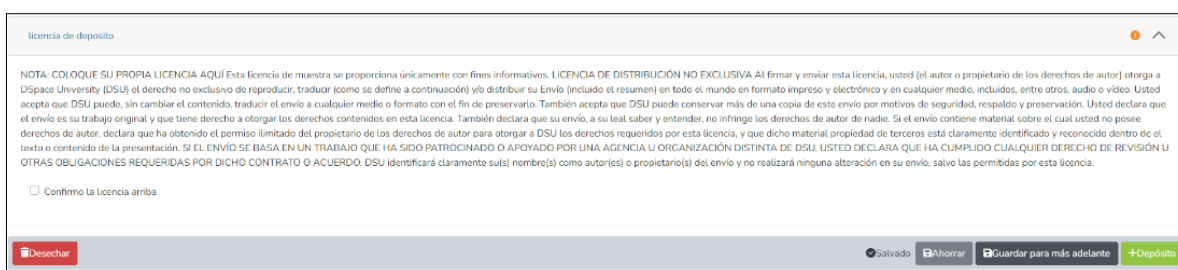


Figura 22. Captura de pantalla de licencia de depósito por defecto en prototipo INVESTUNICAUCA.

La Figura 22 permite visualizar la licencia de términos y condiciones que rige el proceso de depósito en el prototipo. Estos términos establecen las pautas y restricciones para el ingreso de datos de investigación en el repositorio.

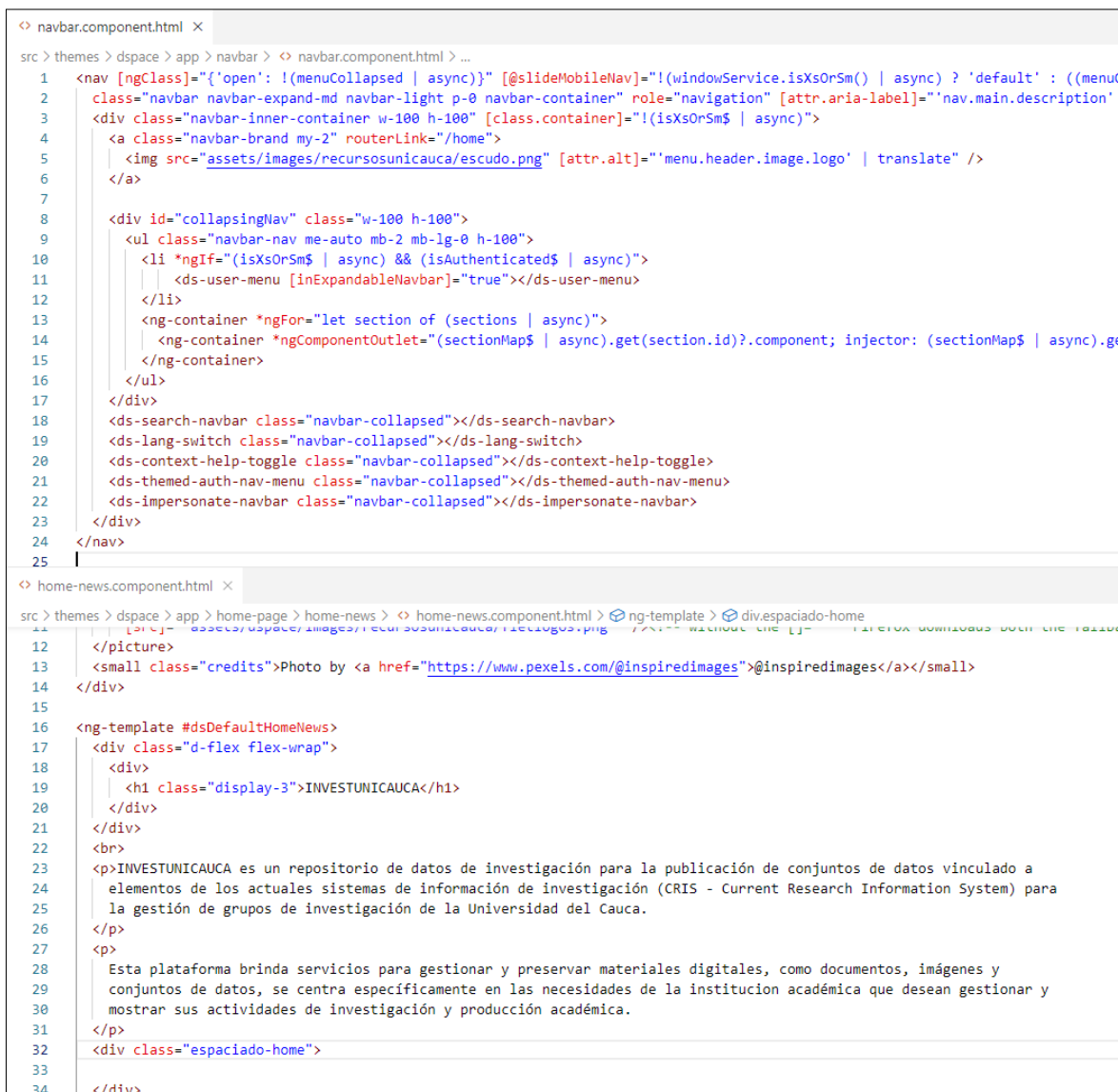
#### 4.6.6 Personalización de interfaz de usuario del prototipo del repositorio de datos de investigación INVESTUNICAUCA

Se realiza la personalización de la interfaz de usuario de la instalación de DSpace-CRIS con base a características de la institución universitaria, teniendo en cuenta colores representativos, bandera, escudos, entre otros también se agregan descripciones sobre el prototipo INVESTUNICAUCA.

Teniendo en cuenta que el aplicativo *web frontend* está construido en angular y que se encuentra organizado por módulos que son bloques de construcción fundamentales que ayudan a organizar y estructurar el código de la aplicación, la personalización entonces se hace en diferentes módulos del sistema en los archivos que lo componen que por lo general son archivo HTML, css y ts, donde el archivo HTML define la estructura de la interfaz de usuario, el archivo css establece los

estilos asociados al componente, y el archivo ts contiene la lógica y funcionalidad del componente.

En la ruta `src\themes\DSpace\app` del código angular se encuentran los módulos como *header*, *header-nav-wrapper*, *home-page*, *navbar*, algunos de los ajustes realizados fueron agregar y cambio de colores representativos de la universidad del cauca en el *navbar*, cambio de imagen el *home-page* y descripción del aplicativo entre otros, a continuación, se muestran algunos de los cambios realizados.



```
1 <nav [ngClass]="{'open': !(menuCollapsed | async)}" [@slideMobileNav]="!(windowService.isXsOrSm() | async) ? 'default' : ((menuC
2 class="navbar navbar-expand-md navbar-light p-0 navbar-container" role="navigation" [attr.aria-label]="nav.main.description'
3 <div class="navbar-inner-container w-100 h-100" [class.container]="!(isXsOrSm$ | async)">
4 <a class="navbar-brand my-2" routerLink="/home">
5 | 
6 </a>
7
8 <div id="collapsingNav" class="w-100 h-100">
9 <ul class="navbar-nav me-auto mb-2 mb-lg-0 h-100">
10 <li *ngIf="(isXsOrSm$ | async) && (isAuthenticated$ | async)">
11 | <ds-user-menu [inExpandableNavbar]="true"></ds-user-menu>
12 </li>
13 <ng-container *ngFor="let section of (sections | async)">
14 | <ng-container *ngComponentOutlet="(sectionMap$ | async).get(section.id).component; injector: (sectionMap$ | async).ge
15 </ng-container>
16 </ul>
17 </div>
18 <ds-search-navbar class="navbar-collapsed"></ds-search-navbar>
19 <ds-lang-switch class="navbar-collapsed"></ds-lang-switch>
20 <ds-context-help-toggle class="navbar-collapsed"></ds-context-help-toggle>
21 <ds-themed-auth-nav-menu class="navbar-collapsed"></ds-themed-auth-nav-menu>
22 <ds-impersonate-navbar class="navbar-collapsed"></ds-impersonate-navbar>
23 </div>
24 </nav>
25
26
27
28
29
30
31
32
33
34
```

```
12 </picture>
13 <small class="credits">Photo by <a href="https://www.pexels.com/@inspiredimages">@inspiredimages</a></small>
14 </div>
15
16 <ng-template #dsDefaultHomeNews>
17 <div class="d-flex flex-wrap">
18 <div>
19 | <h1 class="display-3">INVESTUNICAUCA</h1>
20 </div>
21 </div>
22 <br>
23 <p>INVESTUNICAUCA es un repositorio de datos de investigación para la publicación de conjuntos de datos vinculado a
24 elementos de los actuales sistemas de información de investigación (CRIS - Current Research Information System) para
25 la gestión de grupos de investigación de la Universidad del Cauca.
26 </p>
27 <p>
28 Esta plataforma brinda servicios para gestionar y preservar materiales digitales, como documentos, imágenes y
29 conjuntos de datos, se centra específicamente en las necesidades de la institución académica que desean gestionar y
30 mostrar sus actividades de investigación y producción académica.
31 </p>
32 <div class="espaciado-home">
33
34 </div>
```

Figura 23. Captura de pantalla de código angular personalizado en la interfaz de usuario del prototipo INVESTUNICAUCA.

En seguida, se presenta captura de pantalla de la interfaz de usuario personalizada luego de compilar los cambios realizados en los diferentes módulos de la interfaz de usuario del prototipo INVESTUNICAUCA, el cual se encuentra construido en angular.



Figura 24. Personalización de interfaz de usuario del prototipo INVESTUNICAUCA.

En la Figura 24, se muestra la página principal del prototipo del Repositorio de Datos de Investigación con elementos CRIS (INVESTUNICAUCA). Esta vista inicial proporciona una visión general de la interfaz de usuario y la disposición de elementos diseñados para facilitar la navegación y el acceso a las funcionalidades esenciales del sistema.

#### 4.6.6.1 Aspectos adicionales de configuración para el prototipo del repositorio de datos de investigación INVESTUNICAUCA.

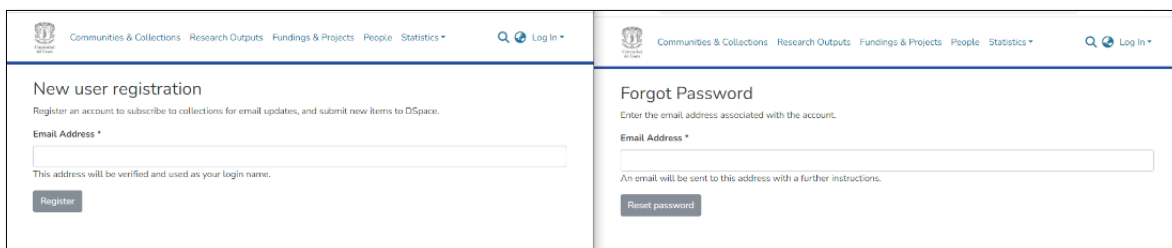
Para el registro de nuevos usuarios y la recuperación de contraseñas de usuarios registrados en el prototipo INVESTUNICAUCA, se hace por medio de correo electrónico, esto se logra con una configuración que se debe definir en *Dspace.cfg*, en este archivo se define el servidor de correo SMTP, la dirección del servidor SMTP que se utilizará para enviar correos electrónicos es Gmail, esto para que se pueda usar las cuentas institucionales de correo electrónico de la universidad del cauca, las credenciales del correo que se utilizará para autenticar el acceso al servidor SMTP, el mismo correo que se utiliza como cuenta *Gmail* de administrador y con *mail.server.port = 465* se especifica el puerto que se utilizará para la comunicación con el servidor SMTP, esta configuración se presenta a continuación:



```
dspace.cfg ×
dspace > config > dspace.cfg
150
151 ##### Email settings #####
152
153 # SMTP mail server (allows DSpace to send email notifications)
154 mail.server = smtp.gmail.com
155
156 # SMTP mail server authentication username and password (if required)
157 mail.server.username = bnarvaez@unicauca.edu.co
158 mail.server.password = xxxxxx
159
160 # SMTP mail server alternate port (defaults to 25)
161 mail.server.port = 465
162
163 # From address for mail
164 # All mail from the DSpace site will use this 'from' address
165 mail.from.address = bnarvaez@unicauca.edu.co
166
167 # Name of a pre-configured Session object to be fetched from a directory.
168 # This overrides the Session settings above. If none can be found, then DSpace
169 # will use the above settings to create a Session.
170 #mail.session.name = Session
171
172 # When feedback is submitted via the Feedback form, it is sent to this address
173 # Currently limited to one recipient!
174 # if this property is empty or commented out, feedback form is disabled
175 feedback.recipient = bnarvaez@unicauca.edu.co
176
177 # General site administration (Webmaster) e-mail
178 # System notifications/reports and other sysadmin emails are sent to this address
179 mail.admin = bnarvaez@unicauca.edu.co
180 mail.admin.name = DSpace Administrator
181
182 # Helpdesk E-mail
183 mail.helpdesk = ${mail.admin}
184 mail.helpdesk.name = Help Desk
185
186 # Recipient for server errors and alerts (defaults to mail.admin)
187 alert.recipient = ${mail.admin}
188
189 # Recipient for new user registration emails (defaults to unspecified)
190 #registration.notify =
191
192 # Enable a "welcome letter" to the newly-registered user.
193 mail.welcome.enabled = false
194
195 # Set the default mail character set. This may be overridden by providing a line
196 # inside the email template "charset: <encoding>", otherwise this default is used.
197 mail.charset = UTF-8
```

Figura 25. Captura de pantalla de configuración servidor SMTP INVESTUNICAUCA.

Una vez realizada la configuración correspondiente a correo electrónico en el prototipo, queda habilitada la creación de usuarios nuevo y la recuperación de contraseña de usuarios registrados, opciones que se encuentran en el *home-page* en la opción de “Log In”

The image shows two side-by-side web forms. The left form is titled "New user registration" and includes a sub-header "Register an account to subscribe to collections for email updates, and submit new items to DSpace." It features an "Email Address \*" input field with a note below it: "This address will be verified and used as your login name." A "Register" button is at the bottom. The right form is titled "Forgot Password" and includes a sub-header "Enter the email address associated with the account." It features an "Email Address \*" input field with a note below it: "An email will be sent to this address with a further instructions." A "Reset password" button is at the bottom. Both forms have a navigation bar at the top with links for "Communities & Collections", "Research Outputs", "Fundings & Projects", "People", and "Statistics", along with a search icon and a "Log In" link.

*Figura 26. Captura de pantalla de creación de nuevo usuario y recuperación de contraseña en el prototipo INVESTUNICAUCA.*

En la Figura 26 se observa la interfaz gráfica destinada a la incorporación de un nuevo usuario, proporcionando un acceso directo y sencillo para la inclusión de información pertinente. Asimismo, se visualiza la interfaz destinada a la recuperación de contraseña, brindando una herramienta eficaz para aquellos usuarios que necesiten restablecer sus credenciales.

#### **4.6.6.2 Roles y grupos en el prototipo del repositorio de datos de investigación INVESTUNICAUCA**

Para interactuar en el sistema una vez se hace un registro de usuario nuevo se deben gestionar por parte de administradores lo que se denomina grupos o roles, donde se establece determinado rol a un usuario registrado, cada rol cuenta con determinados permisos o la posibilidad de realizar ciertas acciones en el sistema, estos grupos o roles se establecen ya sea de manera general en el aplicativo, por comunidades o por colecciones.

De manera general en el aplicativo se tienen unos roles predefinidos, los cuales son Anónimo y Administrador. Mediante la interfaz gráfica se pueden agregar usuarios registrados a estos grupos, el grupo administrador tiene todos los permisos en el aplicativo y el grupo Anónimo es el rol con permisos de solo lectura, en esta interfaz se muestran todos los grupos creados en el aplicativo y se muestra a continuación.

Hogar • Grupos

## Grupos/roles

+ Añadir grupo

### Grupos de búsqueda

Buscar grupos...

Mostrando ahora 1 - 5 de 9

IDENTIFICACIÓN	Nombre	Colección/Comunidad	Miembros	Editar
910706b8-0475-4781-869e-223f5150776b	Anónimo		0	
99d68f9f-ec62-4876-bea0-3b472370b43f	Administrador		3	
1ea56951-fb8a-4ab2-a6d6-8cc1497ee1b5	COMUNIDAD_c8aa51dd-2418-4114-9314-b65c1cf27948_ADMIN	GICO   Grupo de Investigación en Inteligencia Computacional	0	
ff55638d-e426-4d10-9051-142f2367fee6	COLECCIÓN_9775fb16-cd2b-4be6-9a19-15da65de7fc5_SUBMIT	Evento	0	
449f2a1f-e9ea-434f-843d-3ff608c18087	COLECCIÓN_74937578-9225-4a27-89de-ee558eb663f4_SUBMIT	Evento-IDIS	0	

« 1 2 »

Figura 27. Captura de pantalla de interfaz gráfica para grupos y roles en el prototipo INVESTUNICAUCA.

Los grupos que corresponden a una colección o una comunidad en específico son creados en la opción de editar ya sea comunidad o colección, donde aparece la opción “asignar roles”, en esta opción se permite crear grupos para la colección en específico, los grupos de roles que se encuentran disponibles para la creación son los siguientes:

### Roles

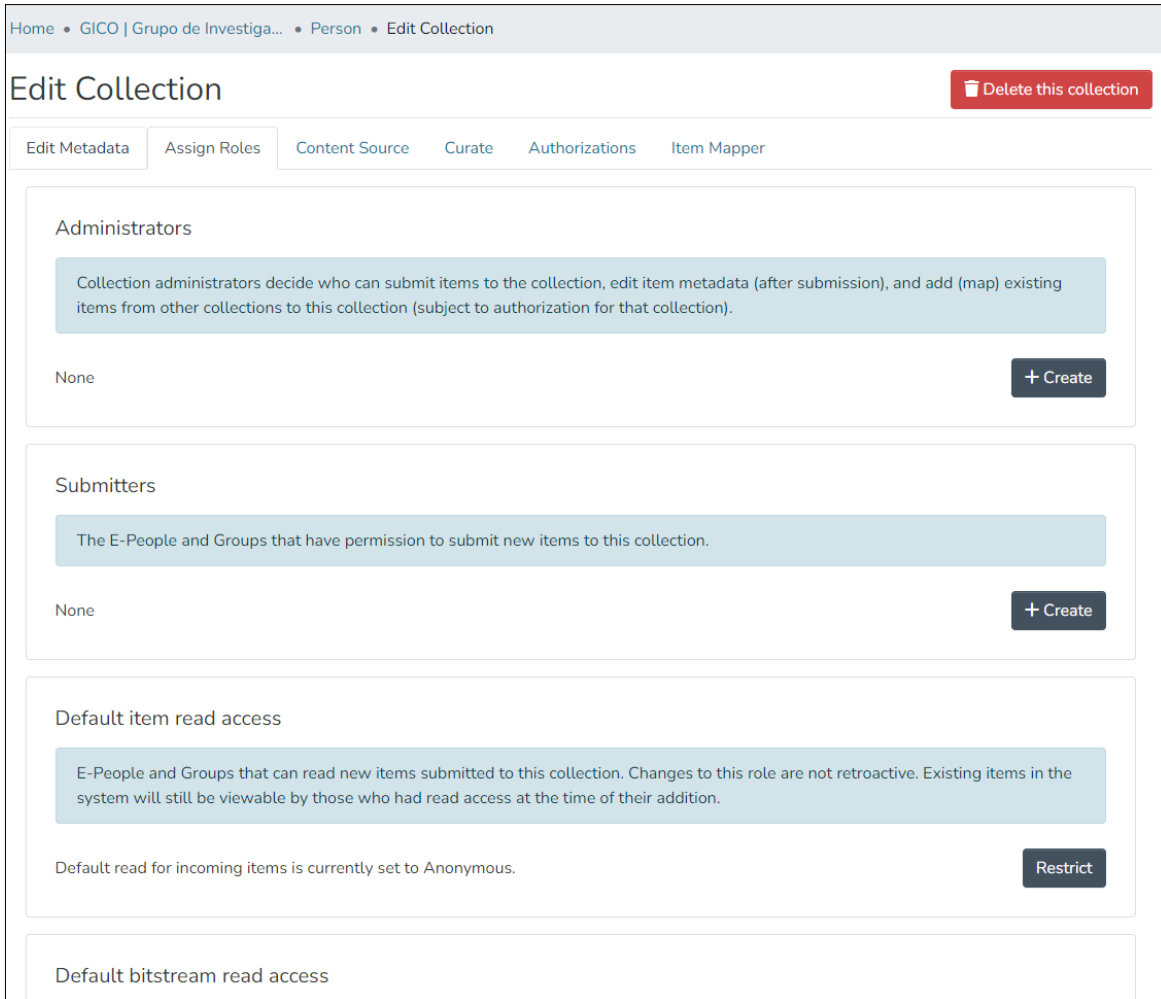
- **Roles administradores:** Los administradores de la colección deciden quién puede enviar elementos a la colección, editar los metadatos de los elementos (después del envío) y agregar (mapear) elementos existentes de otras colecciones a esta colección (sujeto a la autorización para esa colección).
- **Roles presentadores:** Las E-Personas y Grupos que tienen permiso para enviar nuevos elementos a esta colección.

- **Roles revisores:** Los revisores pueden aceptar o rechazar envíos entrantes. Sin embargo, no pueden editar los metadatos del envío.
- **Roles editores:** Los editores pueden editar los metadatos de los envíos entrantes y luego aceptarlos o rechazarlos.
- **Roles editores finales:** Los editores finales podrán editar los metadatos de los envíos entrantes, pero no podrán rechazarlos.

## Accesos

- **Acceso de lectura de elementos predeterminado:** E-Personas y grupos que pueden leer nuevos elementos enviados a esta colección. Los cambios a este rol no son retroactivos. Los elementos existentes en el sistema seguirán siendo visibles para aquellos que tenían acceso de lectura en el momento de su inclusión.
- **Acceso de lectura de flujo de bits predeterminado:** Los administradores de la comunidad pueden crear subcomunidades o colecciones y administrar o asignar administración para esas subcomunidades o colecciones. Además, deciden quién puede enviar elementos a cualquier subcolección, editar los metadatos de los elementos (después del envío) y agregar (mapear) elementos existentes de otras colecciones (sujeto a autorización).

En lo que respecta a asignar roles en una colección en específico, cabe resaltar que algunos de estos grupos se permiten crear y otros solo se pueden restringir.



*Figura 28. Captura de pantalla de interfaz gráfica para asignar roles a una colección en el prototipo INVESTUNICAUCA.*

En la Figura 28, se pueden observar las distintas funciones y responsabilidades que se pueden asignar a los usuarios, permitiendo una gestión precisa y adaptada a las necesidades de la colección en cuestión. La capacidad de definir roles específicos potencia la flexibilidad y la personalización en la administración de la colección, asegurando que cada usuario desempeñe su función de acuerdo con los permisos y responsabilidades asignados.

## **4.7 Revisión de funcionalidades de gestión de un ítem en el prototipo**

### **INVESTUNICAUCA.**

En la siguiente sección, se presentan los casos diseñados para la revisión (de software) al prototipo del repositorio de datos de investigación, INVESTUNICAUCA. Cada caso se ha estructurado para la revisión del prototipo, estas validaciones se centran en evaluar las funciones y comportamientos específicos del prototipo con el objetivo de asegurar que funcionen como se espera.

#### **4.7.1 Definición de casos diseñados para la revisión de software.**

En un enfoque adicional para asegurar la calidad y eficiencia del prototipo, se llevarán a cabo un conjunto de revisiones de las funcionalidades y conformidad con el modelo propuesto.

En este apartado se describen en detalle cada una de las revisiones que se identificaron como necesarias para verificar la gestión de un ítem en el prototipo, estas tres funcionalidades son: la creación, edición y la eliminación de un ítem por parte de un usuario en el sistema, teniendo en cuenta que el usuario ya tiene los roles y permisos respectivos y que en el aplicativo ya se encuentran creada al menos una comunidad y la colección que contiene o contendrá el ítem.

En la Tabla 18 se define la revisión de la funcionalidad de creación de un ítem en una colección del tipo “*Dataset or other products*”, para la creación de este ítem cabe resaltar que, de manera opcional, puede estar asociado a un ítem de proyecto y a un ítem de persona, que sería el autor o los autores del ítem de “*Dataset or other products*”, estos ítems asociados deben estar previamente creados.

Tabla 18. Revisión de funcionalidad agregar ítem en colección

Agregar ítem en colección	CR01
<p>Descripción: Creación de un ítem en la colección “Dataset or other products” del prototipo INVESTUNICAUCA.</p>	
<p>Prerrequisitos</p> <ul style="list-style-type: none"> <li>• Se debe haber iniciado sesión en el aplicativo</li> <li>• Debe existir la comunidad que contiene la colección</li> <li>• Debe existir la colección donde se va a crear el ítem</li> <li>• Se debe contar con un rol que permita crear ítems en el aplicativo</li> </ul>	
<p>Pasos:</p> <ul style="list-style-type: none"> <li>• Seleccionar la opción "Gestión" en el menú principal.</li> <li>• Se selecciona la opción “New”</li> <li>• Se selecciona la opción ítem</li> <li>• Se llena el formulario del ítem a crear</li> <li>• Se debe agregar el archivo adjunto al ítem</li> <li>• Se confirma la licencia de deposito</li> <li>• Se selecciona la opción depositar</li> </ul>	
<p>Resultado esperado:</p> <ul style="list-style-type: none"> <li>• El sistema debería informar "Ítem agregado" y mostrar la vista del ítem agregado en el sistema.</li> <li>• La vista del ítem debería incluir [detalles específicos, como título, autor, fecha, etc.].</li> </ul>	
<p>Resultado obtenido:</p> <ul style="list-style-type: none"> <li>• El sistema informa "Ítem agregado" y muestra la vista del ítem agregado en el sistema.</li> </ul>	

En la Tabla 19 se define la revisión de la funcionalidad de eliminación de un ítem en una colección del tipo “*Dataset or other products*”, se debe tener en cuenta que, para encontrar la opción de la eliminación de un ítem, esta se encuentra inmersa entre la vista de edición, de este modo se debe seleccionar la opción “Editar”, seleccionar el ítem a eliminar y se podrá encontrar la opción de eliminado permanente entre las opciones de “estado”.

Tabla 19. Revisión de funcionalidad eliminar ítem en colección

Eliminar ítem en colección	CR02
<p>Descripción:            Eliminar un ítem de la colección “Dataset or other products” del prototipo INVESTUNICAUCA</p>	
<p>Prerrequisitos</p> <ul style="list-style-type: none"> <li>• Se debe haber iniciado sesión en el aplicativo</li> <li>• Debe existir la comunidad que contiene la colección</li> <li>• Debe existir la colección donde se va a eliminar el ítem</li> <li>• Se debe contar con un rol que permita eliminar ítems en el aplicativo</li> <li>• Debe existir el ítem a eliminar</li> </ul>	
<p>Pasos:</p> <ul style="list-style-type: none"> <li>• Seleccionar la opción "Gestión" en el menú principal.</li> <li>• Se selecciona la opción “Editar”</li> <li>• Se selecciona el ítem a eliminar</li> <li>• Se debe estar en la opción de estado</li> <li>• Seleccionar la opción “Borrar permanentemente”</li> <li>• Confirmar la eliminación del ítem en la opción “Borrar”.</li> </ul>	
<p>Resultado esperado:</p> <ul style="list-style-type: none"> <li>• El sistema debería informar "Ítem eliminado" y confirmar la exitosa eliminación del ítem.</li> <li>• El ítem eliminado no debería ser visible en la colección " Dataset or other products ".</li> </ul>	
<p>Resultado obtenido:</p> <ul style="list-style-type: none"> <li>• El sistema informa "Ítem eliminado" y confirma la exitosa eliminación del ítem.</li> <li>• El ítem eliminado no es visible en la colección " Dataset or other products ".</li> </ul>	



En la Tabla 20 se define el caso de revisión para la edición de un ítem en una colección del tipo “*Dataset or other products*”, para la edición de este ítem existe una opción de editar ítem, la cual brinda una vista con una serie de opciones para editar como el estado, metadatos, historial de versión y flujos de bits, esta revisión se enfoca en la edición de metadatos del ítem.

Tabla 20. Revisión de funcionalidad editar ítem en colección

Editar ítem en colección	CR03
<p>Descripción: Edición de un ítem de la colección “Dataset or other products” del prototipo INVESTUNICAUCA</p>	
<p>Prerrequisitos</p> <ul style="list-style-type: none"> <li>• Se debe haber iniciado sesión en el aplicativo.</li> <li>• Debe existir la comunidad que contiene la colección.</li> <li>• Debe existir la colección donde se va a editar el ítem.</li> <li>• Se debe contar con un rol que permita editar ítems en el aplicativo.</li> <li>• Debe existir el ítem a editar en la colección.</li> </ul>	
<p>Pasos:</p> <ul style="list-style-type: none"> <li>• Seleccionar la opción "Gestión" en el menú principal.</li> <li>• Se selecciona la opción “Editar”</li> <li>• Se selecciona el ítem a editar</li> <li>• Se debe estar en la opción de metadatos</li> <li>• Se modifican los metadatos que se desean editar.</li> <li>• Se selecciona el botón “Ahorrar”</li> </ul>	
<p>Resultado esperado:</p> <ul style="list-style-type: none"> <li>• El sistema debería informar "Metadatos guardados".</li> <li>• Los cambios realizados en el ítem (en los campos especificados) deberían reflejarse correctamente.</li> </ul>	
<p>Resultado obtenido:</p> <ul style="list-style-type: none"> <li>• El sistema informa " Metadatos guardados ".</li> <li>• Los cambios realizados en el ítem se reflejan correctamente.</li> </ul>	

La ejecución exitosa de estos de los casos de revisión definidos anteriormente respalda la calidad y confiabilidad del prototipo INVESTUNICAUCA en lo que respecta a la gestión de ítems en la colección del tipo “*Dataset or other products*”. Los resultados obtenidos han demostrado coherencia con las expectativas,

confirmando el correcto funcionamiento de las operaciones evaluadas, también estos resultados contribuyen a la entrega de un producto final confiable y a captura de posibles mejoras dado el caso.

Estas revisiones han permitido evaluar la funcionalidad del sistema, también han validado la integración entre DSpace con elementos CRIS. Los resultados obtenidos ofrecen una visión de la viabilidad del enfoque adoptado, consolidando así un paso significativo hacia la mejora de la gestión de datos de investigación en un ámbito universitario. Con el respaldo de estos casos de revisión, se sientan las bases para futuras implementaciones y mejoras.

## 5. CONCLUSIONES

- El mapeo sistemático realizado brinda un panorama actual de las temáticas que comprende esta investigación, el cual trascendió a una revisión sistemática que permitió la definición del modelo de publicación de datos propuesto y destaca: la importancia de los datos de investigación como datos que respaldan los diferentes procesos investigativos, la necesidad de que los datos cumplan los principios FAIR para garantizar su gestión, que los repositorios de datos de investigación son fundamentales para la eficiencia y transparencia en la investigación, la reutilización de datos y la generación de nuevo conocimiento.
- El presente trabajo ha puesto de manifiesto la trascendencia de los metadatos como factor clave, así como el manejo de estándares de metadatos en lo referente a la gestión de datos de investigación, dado que estos son un componente esencial que permiten garantizar la coherencia, la recopilación eficiente de datos, la accesibilidad, y la interoperabilidad en aspectos relacionados a información y datos producto de investigaciones científicas.
- Es importante enfatizar que la adopción de los principios FAIR (*Findable, Accessible, Interoperable, Reusable*) optimizan la visibilidad de los datos en un entorno de investigación y para que los datos cumplan con estos principios es necesario el uso de identificadores persistentes, la adopción de estándares aceptados por la comunidad científica como por ejemplo, Dublín Core, y la adopción de modelos como CERIF para lograr formatos comunes en la representación de datos y descripción detallada de los mismos.
- La integración entre un repositorio de datos y elementos CRIS respaldado por un modelo estructurado, demuestra ser un paso trascendente para la gestión y representación efectiva de datos de investigación en el contexto universitario, dado que brinda una visión completa acerca de los procesos de investigación llevados a cabo por los grupos de investigación en una Universidad y a su vez permite obtener análisis detallados, aumenta la visibilidad de estos procesos, así como también facilita la reutilización de datos de investigación producidos y almacenados. Por lo tanto, este tipo de integración se presenta como una solución adecuada que aborda las necesidades específicas de los diversos grupos o actores involucrados en investigación científica.
- La adaptación de un DSpace CRIS permitió integrar las características definidas en la propuesta de modelo de publicación de datos, y su implementación puede realizarse en diversos grupos de investigación dentro de una universidad u otras instituciones de educación superior, porque se

garantiza su versatilidad, escalabilidad y aplicación generalizada, al considerarse características como: el perfil de aplicación, formatos de archivos aceptados, entidades CRIS, entre otros.

- El prototipo INVESTUNICAUCA representa una herramienta que se enmarca en el modelo colaborativo denominado ciencia abierta y es útil para almacenar, preservar y compartir datos de investigación, integrando características como: la persistencia centralizada y organizada, uso de estándares de metadatos, identificadores persistentes, acceso a recursos considerando licencias y derechos de autor, etc.

## 6. TRABAJOS FUTUROS

- Se empleó el software DSpace CRIS en su versión 7.0, la cual, hasta el momento de la creación del prototipo, no cuenta con compatibilidad para la gestión de licencias Creative Commons. Como posible línea de trabajo futuro, se recomienda validar si en posteriores desarrollos de la comunidad que da soporte a DSpace llega a habilitar el trabajo con estas licencias.
- Como trabajo futuro, se propone llevar a cabo una evaluación integral del impacto del prototipo “INVESTUNICAUCA” en el contexto en el que fue definido en un tiempo determinado. Este análisis podría abarcar distintos aspectos como adopción y usabilidad, retroalimentación de usuarios, calidad de datos, entre otros aspectos que serán clave para entender el alcance y la efectividad del sistema.
- Desarrollar estrategias que permitan garantizar el cumplimiento de buenas prácticas en la gestión de datos de investigación. Este enfoque incluye investigar los estándares emergentes de metadatos que contribuyan a mejorar y potenciar la interoperabilidad, otro aspecto a considerar es la exploración de otras plataformas que sigan siendo robustas y actualizadas las cuales garanticen la eficiencia y transparencia de gestión, así como que se alineen con las necesidades cambiantes de la comunidad académica.

## 7. REFERENCIAS

- Adriaan, D., Smits, B., & Teperek, M. (2020). *Research Data Management for Master 's Students : From Awareness to Action*. Miyakawa, 1–11.
- Aleixandre-benavent, R., Vidal-infer, A., & Alonso-arroyo, A. (2020). *Research Data Sharing in Spain : Exploring*. 1–14.
- Arend, D., Patrick, K., Lange, M., Junker, A., & Scholz, U. (2020). *The on-premise data sharing infrastructure e! DAL : Foster FAIR data for faster data acquisition*. iii, 1–11. <https://doi.org/10.1093/gigascience/giaa107>
- Austin, T., Bei, K., Efthymiadis, T., & Koumoulos, E. P. (2021). *Lessons Learnt from Engineering Science Projects Participating in the Horizon 2020 Open Research Data Pilot*. 1–19.
- Bahim, C., Casorrán-amilburu, C., Dekkers, M., Herczog, E., Loozen, N., Repanas, K., Russell, K., & Stall, S. (2020). *The FAIR Data Maturity Model: An Approach to Harmonise FAIR Assessments*. 1–7.
- Bellgard, M. I. (2020). International Journal of Information Management ERDMAS : An exemplar-driven institutional research data management and analysis strategy. *International Journal of Information Management*, 50(August 2019), 337–340. <https://doi.org/10.1016/j.ijinfomgt.2019.08.009>
- Bezuidenhout, L., Drummond-Curtis, S., Walker, B., Shanahan, H., & Alfaro-Córdoba, M. (2021). A school and a network: Codata-rda data science summer schools alumni survey. *Data Science Journal*, 20(1), 1–12. <https://doi.org/10.5334/DSJ-2021-010>
- Bhardwaj, R. K. (2019). Open research data repositories: A content analysis to comprehend data equitable access. *Journal of Scientometric Research*, 8(3), 135–142. <https://doi.org/10.5530/JSCIRES.8.3.29>
- Biesenbender, S., Petersohn, S., & Thiedig, C. (2019). Using Current Research Information Systems (CRIS) to showcase national and institutional research (potential): Research information systems in the context of Open Science. *Procedia Computer Science*, 146(2018), 142–155. <https://doi.org/10.1016/j.procs.2019.01.089>
- Bloemers, M., & Montesanti, A. (2020). The fair funding model: Providing a framework for research funders to drive the transition toward fair data management and stewardship practices. *Data Intelligence*, 2(1–2), 171–180. [https://doi.org/10.1162/dint\\_a\\_00039](https://doi.org/10.1162/dint_a_00039)
- Borycz, J. (2021). Implementing data management workflows in research groups through integrated library consultancy. *Data Science Journal*, 20(1), 1–9. <https://doi.org/10.5334/DSJ-2021-009>
- Bossaller, J., & Million, A. J. (2022). The research data life cycle, legacy data, and dilemmas in research data management. *Journal of the Association for Information Science and Technology*, August 2021, 1–6. <https://doi.org/10.1002/asi.24645>
- Brandt, N., Garabedian, N. T., Schoof, E., Schreiber, P. J., Zschumme, P., Greiner, C., & Selzer, M. (2022). Managing FAIR Tribological Data Using Kadi4Mat. *Data*, 7(2), 1–16.

<https://doi.org/10.3390/data7020015>

- Brandt, N., Griem, L., Herrmann, C., Schoof, E., Tosato, G., Zhao, Y., Zschumme, P., & Selzer, M. (2021). Kadi4mat: A research data infrastructure for materials science. *Data Science Journal*, 20(1), 1–14. <https://doi.org/10.5334/dsj-2021-008>
- Buchholz, P. (2021). *Adaptable Methods for Training in Research Data Management*. 1–9. <https://doi.org/10.5334/dsj-2021-014>
- Calhoun, S. P., Akin, D., Zimmerman, B., & Neeman, H. (2019). Large scale research data archiving : Training for an inconvenient technology. *Journal of Computational Science*, 36, 100523. <https://doi.org/10.1016/j.jocs.2016.07.005>
- Cerda-Cosme, R., & Méndez, E. (2022). Analysis of shared research data in Spanish scientific papers about COVID-19: A first approach. *Journal of the Association for Information Science and Technology*, January, 1–13. <https://doi.org/10.1002/asi.24716>
- cern. (2023). *About CERN*. Cern. <https://www.home.cern/about>
- Cho, H., & Yu, Y. (2018). *Link prediction for interdisciplinary collaboration via co - authorship network*. 1–12.
- ckan.org. (2023). *El sistema de gestión de datos de código abierto líder en el mundo*. Ckan.Org. <https://ckan.org/>
- Commission, E. (2020). *H2020 Programme Guidelines to the Rules on Open Access to Scientific Publications and Open Access to Research Data in Horizon 2020*. March 2017.
- Commons, C. (n.d.). *Sobre las licencias - Creative Commons*. Retrieved July 4, 2023, from <https://creativecommons.org/licenses/?lang=es>
- Cousijn, H., Feeney, P., Lowenberg, D., Presani, E., & Simons, N. (2019). Bringing citations and usage metrics together to make data count. *Data Science Journal*, 18(1), 1–7. <https://doi.org/10.5334/dsj-2019-009>
- Cox, A. M., Kennan, M. A., Lyon, L., Pinfield, S., & Sbaffi, L. (2019). Maturing research data services and the transformation of academic libraries. In *Journal of Documentation* (Vol. 75, Issue 6). <https://doi.org/10.1108/JD-12-2018-0211>
- Cox, A. M., Wan, W., Tam, T., Cox, A. M., Wan, W., & Tam, T. (2018). *A critical analysis of lifecycle models of the research process and research data management management*. <https://doi.org/10.1108/AJIM-11-2017-0251>
- Cruz, M., Dintzner, N., Dunning, A., Kuil, A. van der, Plomp, E., Teperek, M., Velden, Y. T. van der, & Versteeg, A. (2019). Policy needs to go hand in hand with practice: The learning and listening approach to data management. *Data Science Journal*, 18(1), 1–11. <https://doi.org/10.5334/dsj-2019-045>
- Curdt, C., Dierkes, J., & Kloppenburg, S. (2022). RDM in a Decentralised University Ecosystem—A Case Study of the University of Cologne. *Data Science Journal*, 21(1), 1–8. <https://doi.org/10.5334/dsj-2022-020>

- DataCite. (n.d.). *Welcome to DataCite*. Retrieved May 20, 2023, from <https://datacite.org/>
- David, R., Mabile, L., Specht, A., Stryeck, S., Thomsen, M., Yahia, M., Jonquet, C., Dollé, L., Jacob, D., Bailo, D., Bravo, E., Gachet, S., Gunderman, H., Hollebecq, J. E., Ioannidis, V., Bras, Y. Le, Lerigoleur, E., & Cambon-Thomsen, A. (2020). Fairness literacy: The achilles' heel of applying fair principles. *Data Science Journal*, *19*(1), 1–11. <https://doi.org/10.5334/dsj-2020-032>
- DDI Alliance. (n.d.). *Welcome to the Data Documentation Initiative | Data Documentation Initiative*. Retrieved May 20, 2023, from <https://ddialliance.org/>
- de la Hidalga, A. N., Decarolis, D., Xu, S., Matam, S., Enciso, W. Y. H., Goodall, J., Matthews, B., & Catlow, C. R. A. (2022). A Workflow Demonstrator for Processing Catalysis Research Data. *Data Intelligence*, *4*(2), 455–470. [https://doi.org/10.1162/dint\\_a\\_00143](https://doi.org/10.1162/dint_a_00143)
- Defranco, J. F., & Voas, J. (2021). Reproducibility, Fabrication, and Falsification. *Computer*, *54*(12), 24–26. <https://doi.org/10.1109/MC.2021.3055926>
- Downs, R. R. (2021). *Improving Opportunities for New Value of Open Data : Assessing and Certifying Research Data Repositories*. 1–11. <https://doi.org/10.5334/dsj-2021-001>
- DSpace-CRIS. (2023). *DSpace-CRIS Home - DSpace-CRIS - LYRISIS Wiki*. <https://wiki.lyrasis.org/display/DSPACECRIS/DSpace-CRIS+Home>
- DSpace.lyrasis.org. (2023). *About DSpace*. DSpace.Lyrasis.Org. <https://DSpace.lyrasis.org/about/>
- Dublin Core. (n.d.). *DCMI: Home*. Retrieved May 17, 2023, from <https://www.dublincore.org/>
- Dvořák, J., Dvořák, J., Chudlarský, T., & Špaček, J. (2019). ScienceDirect Procedia ScienceDirect Practical CRIS Interoperability Practical. *Procedia Computer Science*, *146*, 256–264. <https://doi.org/10.1016/j.procs.2019.01.077>
- Ekström, B. (2023). Thousands of examining eyes: credibility, authority and validity in biodiversity citizen science data production. *Aslib Journal of Information Management*, *75*(1), 149–170. <https://doi.org/10.1108/AJIM-10-2021-0292>
- Elsevier. (n.d.-a). *What is Scopus Preview? - Scopus: Access and use Support Center*. Retrieved June 8, 2023, from [https://service.elsevier.com/app/answers/detail/a\\_id/15534/supporthub/scopus/#tips](https://service.elsevier.com/app/answers/detail/a_id/15534/supporthub/scopus/#tips)
- Elsevier. (n.d.-b). *Why choose Scopus - Scopus benefits | Elsevier solutions*. Retrieved June 19, 2023, from <https://www.elsevier.com/solutions/scopus/why-choose-scopus>
- Elsevier. (2020). *Content Coverage Guide. Scopus*, 1–24. [https://www.elsevier.com/\\_data/assets/pdf\\_file/0007/69451/Scopus\\_ContentCoverage\\_Guide\\_WEB.pdf](https://www.elsevier.com/_data/assets/pdf_file/0007/69451/Scopus_ContentCoverage_Guide_WEB.pdf) (Accessed on 22 July 2021)
- Enescu, I. I., Plattner, G. K., Pernas, L. E., Haas-Artho, D., Bischof, S., Lehning, M., & Steffen, K. (2018). The EnviDat concept for an institutional environmental data portal. *Data Science Journal*, *17*, 1–17. <https://doi.org/10.5334/dsj-2018-028>



- EuroCRIS. (n.d.). *GitHub - EuroCRIS/CERIF-Core: The CERIF Refactoring Pilot project started in 2021 by euroCRIS*. Retrieved December 6, 2023, from <https://github.com/EuroCRIS/CERIF-Core>
- EuroCRIS. (2020). *Main features of CERIF*. <https://eurocris.org/services/main-features-cerif>
- Fabre, R., & Egret, D. (2012). *Evaluating the scientific impact of research infrastructures: The role of current research information systems*. <https://doi.org/10.1162/qss>
- Farinelli, C., & Zigoni, A. (2022). Extending the value of a CRIS with Research Data Management. *Procedia Computer Science*, 211(C), 187–195. <https://doi.org/10.1016/j.procs.2022.10.190>
- fedora.lyrasis.org. (2023). *About Us*. Fedora.Lyrasis.Org. <https://fedora.lyrasis.org/about/>
- figshare. (2023). *the world of figshare*. Figshare. <https://knowledge.figshare.com/>
- Fitschen, T., Schlemmer, A., Hornung, D., Tom Wörden, H., Parlitz, U., & Luther, S. (2019). *CaosDB—research data management for complex, changing, and automated research workflows*. *Data*, 4(2), 1–11. <https://doi.org/10.3390/data4020083>
- Gaignard, A., Rosnet, T., Lamotte, F. De, Lefort, V., & Devignes, M. D. (2023). FAIR - Checker : supporting digital resource findability and reuse with Knowledge Graphs and Semantic Web standards. *Journal of Biomedical Semantics*, 1–14. <https://doi.org/10.1186/s13326-023-00289-5>
- Ganske, A., Heydebreck, D., Höck, H., Kraft, A., Quaas, J., & Kaiser, A. (2020). *A short guide to increase FAIRness of atmospheric model data*. 29(6), 483–491. <https://doi.org/10.1127/metz/2020/1042>
- GICO | Grupo de Investigación en Inteligencia Computacional. (n.d.). Retrieved September 14, 2023, from <https://www.unicauca.edu.co/gico/>
- Gonzales, S., Carson, M. B., & Holmes, K. (2022). Ten simple rules for maximizing the recommendations of the NIH data management and sharing plan. *PLoS Computational Biology*, 18(8), 1–16. <https://doi.org/10.1371/journal.pcbi.1010397>
- Graf, C., Flanagan, D., Wylie, L., & Silver, D. (2020). The open data challenge: An analysis of 124,000 data availability statements and an ironic lesson about data management plans. *Data Intelligence*, 2(4), 554–568. [https://doi.org/10.1162/dint\\_a\\_00061](https://doi.org/10.1162/dint_a_00061)
- Grasse, M., López, A., & Winter, N. (2018). Landesinitiative NFDI – A central point of contact for RDM for higher education institutions in the German state of north Rhine-Westphalia. *Data Science Journal*, 17, 1–7. <https://doi.org/10.5334/dsj-2018-025>
- Gregory, K., Khalsa, S. J., Michener, W. K., Psomopoulos, F. E., Waard, A. De, & Wu, M. (2018). *Eleven quick tips for finding research data*. 3–9.
- Gregory, K. M., Cousijn, H., Groth, P., Scharnhorst, A., & Wyatt, S. (2020). *Understanding data search as a socio-technical practice*. <https://doi.org/10.1177/0165551519837182>
- Grunzke, R., Hartmann, V., Jejkal, T., Kollai, H., Prabhune, A., Herold, H., Deicke, A.,

- Dressler, C., Dolhoff, J., Stanek, J., Hoffmann, A., Müller-pfefferkorn, R., Schrade, T., Meinel, G., Herres-pawlis, S., & Nagel, W. E. (2019). The MASi repository service — Comprehensive , metadata-driven and multi-community research data management. *Future Generation Computer Systems*, 94, 879–894. <https://doi.org/10.1016/j.future.2017.12.023>
- Hansson, K., & Dahlgren, A. (2022). *Open research data repositories : Practices , norms , and metadata for sharing images*. August 2021, 303–316. <https://doi.org/10.1002/asi.24571>
- Hashim, H. N. M. (2019). Developing a model guidelines addressing legal impediments to open access to publicly funded research data in Malaysia. *Data Science Journal*, 18(1). <https://doi.org/10.5334/dsj-2019-027>
- Hasselbring, W., Carr, L., Hettrick, S., Packer, H., & Tiropanis, T. (2020). From FAIR research data toward FAIR and open research software. *IT - Information Technology*, 62(1), 39–47. <https://doi.org/10.1515/itit-2019-0040>
- Hedeland, H. (2020). Providing digital infrastructure for audio-visual linguistic research data with diverse usage scenarios: Lessons learnt. *Publications*, 8(2). <https://doi.org/10.3390/PUBLICATIONS8020033>
- Hettne, K. M., Verhaar, P., Schultes, E., & Sesink, L. (2020). *From FAIR Leading Practices to FAIR Implementation and Back : An Inclusive Approach to FAIR at Leiden University Libraries*. *CoreTrustSeal 2019*, 1–7.
- Hrynaszkiewicz, I., Simons, N., Hussain, A., Grant, R., & Goudie, S. (2020). *Developing a Research Data Policy Framework for All Journals and Publishers*. 1–15.
- Iglezakis, D. (2020). *EngMeta – Metadata for Computational Engineering*. 14(1), 26–38.
- inveniosoftware.org. (2023). *About Invenio*. Inveniosoftware.Org.
- Ishak, I., Tajuddin, N. I. I., Yah Jusoh, Y., Sidi, F., Abdullah, R., Marhaban, H., Tugiran, Y., & Yusof, Y. (2019). Influencing factors in determining research data repository infrastructure for research data management. *International Journal of Engineering and Advanced Technology*, 9(1), 1655–1660. <https://doi.org/10.35940/ijeat.A2645.109119>
- Islam, S., Hardisty, A., Addink, W., Weiland, C., & Glöckler, F. (2020). *Incorporating RDA Outputs in the Design of a European Research Infrastructure for Natural Science Collections*. 1–14.
- Jackson, B. (2021). Open data policies among library and information science journals. *Publications*, 9(2). <https://doi.org/10.3390/publications9020025>
- Jejkal, T., Chelbi, S., Pfeil, A., & Wittenburg, P. (2022). *Evaluation of Application Possibilities for Packaging Technologies in Canonical Workflows*. 0–3. <https://doi.org/10.1162/dint>
- Jetten, M., Simons, E., & Rijnders, J. (2019). The role of CRIS's in the research life cycle. A case study on implementing a FAIR RDM policy at Radboud University, the Netherlands. *Procedia Computer Science*, 146, 156–164. <https://doi.org/10.1016/j.procs.2019.01.090>

- Jiao, C., & Li, K. (2022). *Data sharing practices across knowledge domains : A dynamic examination of data availability statements in PLOS ONE publications*. <https://doi.org/10.1177/01655515221101830>
- Kern, D., & Hienert, D. (2018). Understanding the information needs of social scientists in Germany. *Proceedings of the Association for Information Science and Technology*, 55(1), 234–243. <https://doi.org/10.1002/pra2.2018.14505501026>
- Kindling, M., & Strecker, D. (2022). Data Quality Assurance at Research Data Repositories. *Data Science Journal*, 21(1), 1–17. <https://doi.org/10.5334/dsj-2022-018>
- Klump, J., Fleischer, D., Ramdeen, S., & Wyborn, L. (2021). *Towards Globally Unique Identification of Physical Samples : Governance and Technical Implementation of the IGSN Global Sample Number*. 1–16. <https://doi.org/10.5334/dsj-2021-033>
- Koltay, T. (2020). Quality of open research data: Values, convergences and governance. *Information (Switzerland)*, 11(4). <https://doi.org/10.3390/info11040175>
- Korkeamäki, L., Keskustalo, H., & Kumpulainen, S. (2022). Task information types related to data gathering in media studies. *Journal of Documentation*, 78(7), 528–545. <https://doi.org/10.1108/JD-04-2022-0082>
- Labastida, I., & Margoni, T. (2020). *Licensing FAIR Data for Reuse*. 5, 0–1. <https://doi.org/10.1162/dint>
- Latif, A., Limani, F., & Tochtermann, K. (2019). A generic research data infrastructure for long tail research data management. *Data Science Journal*, 18(1), 1–11. <https://doi.org/10.5334/dsj-2019-017>
- Latif, A., Limani, F., & Tochtermann, K. (2021). *On the Complexities of Federating Research Data Infrastructures*. 0–3. <https://doi.org/10.1162/dint>
- Lefebvre, A. (2023). *Laboratory Forensics for Open Science Readiness : an Investigative Approach to Research Data Management*. 381–399.
- Lefebvre, A., Bakhtiari, B., & Spruit, M. (2020). Exploring research data management planning challenges in practice. *IT - Information Technology*, 62(1), 29–37. <https://doi.org/10.1515/itit-2019-0029>
- Lehmann, J., Schorz, S., Rache, A., Häußermann, T., Rädle, M., & Reichwald, J. (2023). Establishing Reliable Research Data Management by Integrating Measurement Devices Utilizing Intelligent Digital Twins. *Sensors*, 23(1). <https://doi.org/10.3390/s23010468>
- LYRASIS. (2011). *Functional Overview*. <https://wiki.lyrasis.org/display/DSDOC17/Functional+Overview>
- LYRASIS. (2017). *Management of CRIS entities - DSpace-CRIS - LYRASIS Wiki*. <https://wiki.lyrasis.org/display/DSPACECRIS/Management+of+CRIS+entities>
- Mahony, S. (2022). *Toward openness and transparency to better facilitate knowledge creation*. *July 2021*, 1474–1488. <https://doi.org/10.1002/asi.24652>

- Manik, L. P., Akbar, Z., & Yaman, A. (2022). *Indonesian Scientists' Behavior Relative to Research Data Governance in Preventing WMD-Applicable Technology Transfer*.
- Mannheimer, S., Clark, J. A., Espeland, J., & Hagerman, K. (2019). *Building a Dataset Search for Institutions : Project Update*.
- Marlina, E., Hidayanto, A. N., & Purwandari, B. (2022). Towards a model of research data management readiness in Indonesian context: An investigation of factors and indicators through the fuzzy delphi method. *Library and Information Science Research*, 44(1), 101141. <https://doi.org/10.1016/j.lisr.2022.101141>
- Masinde, J., Chen, J., Wambiri, D., & Mumo, A. (2021). Research Librarians' Experiences of Research Data Management Activities at an Academic Library in a Developing Country. *Data and Information Management*, 5(4), 412–424. <https://doi.org/10.2478/dim-2021-0002>
- McKeague, P., Corns, A., Larsson, Å., Moreau, A., Posluschny, A., Van Daele, K., & Evans, T. (2020). One Archaeology: A manifesto for the systematic and effective use of mapped data from archaeological fieldwork and research. *Information (Switzerland)*, 11(4). <https://doi.org/10.3390/INFO11040222>
- Medina-Smith, A., Becker, C. A., Plante, R. L., Bartolo, L. M., Dima, A., Warren, J. A., & Hanisch, R. J. (2021). A controlled vocabulary and metadata schema for materials science data discovery. *Data Science Journal*, 20(1), 1–10. <https://doi.org/10.5334/dsj-2021-018>
- Miksa, T., Walk, P., Neish, P., Oblasser, S., Murray, H., Renner, T., Jacquemot-Perbal, M. C., Cardoso, J., Kvamme, T., Praetzellis, M., Suchánek, M., Hooft, R., Faure, B., Moa, H., Hasan, A., & Jones, S. (2021). Application profile for machine-actionable data management plans. *Data Science Journal*, 20(1), 1–17. <https://doi.org/10.5334/DSJ-2021-032>
- Minamiyama, Y., Ikeuchi, U., Ueshima, K., Okayama, N., & Takeda, H. (2020). Investigation and development of the workflow to clarify conditions of use for research data publishing in japan. *Data Science Journal*, 19(1), 1–20. <https://doi.org/10.5334/dsj-2020-053>
- Mushi, G. E., Pienaar, H., & Deventer, M. van. (2020). Identifying and implementing relevant research data management services for the library at the university of dodoma, Tanzania. *Data Science Journal*, 19(1), 1–9. <https://doi.org/10.5334/dsj-2020-001>
- Nie, H., Luo, P., & Fu, P. (2021). Research data management implementation at Peking university library: Foster and promote open science and open data. *Data Intelligence*, 3(1), 189–204. [https://doi.org/10.1162/dint\\_a\\_00088](https://doi.org/10.1162/dint_a_00088)
- Nishida, E., Ishita, E., Watanabe, Y., & Tomiura, Y. (2020). Description of research data in laboratory notebooks: Challenges and opportunities. *Proceedings of the Association for Information Science and Technology*, 57(1), 3–5. <https://doi.org/10.1002/pr2.388>
- Nitecki, D. A., & Alter, A. (2021). Leading fair adoption across the institution: A collaboration between an academic library and a technology provider. *Data Science Journal*, 20(1),

1–8. <https://doi.org/10.5334/dsj-2021-006>

Oliver, Y. L. H., Makoto, C., Wu, M., & Huvila, I. (2020). *Supporting open research data practice through data curation and discovery: A global perspective*. 2–4. <https://doi.org/10.1002/pra2.291>

Pereira, M., César, F., & Leite, L. (2019). *Factors influencing research data communication on Zika virus : a grounded theory*. <https://doi.org/10.1108/JD-05-2018-0071>

Poblet, M., Aryani, A., Manghi, P., Unsworth, K., Wang, J., Hausstein, B., Dallmeier-Tiessen, S., Klas, C. P., Casanovas, P., & Rodriguez-Doncel, V. (2018). Assigning Creative Commons Licenses to Research Metadata: Issues and Cases. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (Vol. 10791). Springer International Publishing. [https://doi.org/10.1007/978-3-030-00178-0\\_16](https://doi.org/10.1007/978-3-030-00178-0_16)

PolicyLab, Departamento Administrativo de Presidencia de la República, Departamento Nacional de Planeación, & Ministerio de Tecnologías de Información y Comunicaciones. (2022). *Modelo de Gobernanza de la Infraestructura de Datos para el Estado Colombiano*.

Posavec, K., Celjak, D., & Musap, L. J. (2020). *Role of a Croatian National Repository Infrastructure in Promotion and Support of Research Data Management*. 1–7.

Pronk, T. E. (2019). The time efficiency gain in sharing and reuse of research data. *Data Science Journal*, 18(1), 6–13. <https://doi.org/10.5334/dsj-2019-010>

Quarati, A., & Raffaghelli, J. E. (2022). Do researchers use open research data? Exploring the relationships between usage trends and metadata quality across scientific disciplines from the Figshare case. *Journal of Information Science*, 48(4), 423–448. <https://doi.org/10.1177/0165551520961048>

Rodrigues Rufino Campêlo, L. R., & Barreto Neto, V. C. (2019). Comparando softwares gratuitos para criação de repositórios de dados abertos. *Ci.Inf., Brasília*, 48(3), 341–346.

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

Rousi, A. M. (2022). *Using current research information systems to investigate data acquisition and data sharing practices of computer scientists*. <https://doi.org/10.1177/09610006221093049>

Rousi, A. M., & Laakso, M. (2020). Journal research data sharing policies: a study of highly-cited journals in neuroscience, physics, and operations research. *Scientometrics*, 124(1), 131–152. <https://doi.org/10.1007/s11192-020-03467-9>

Rumble, J., Broome, J., & Hodson, S. (2019). Building an international consensus on multi-disciplinary metadata standards: A CODATA case history in nanotechnology. *Data Science Journal*, 18(1), 1–11. <https://doi.org/10.5334/dsj-2019-012>

- Schirrwagen, J., Cimiano, P., Ayer, V., Pietsch, C., Wiljes, C., Vompras, J., & Pieper, D. (2019). Expanding the research data management service portfolio at bielefeld university according to the three-pillar principle towards data FAIRness. *Data Science Journal*, 18(1), 1–10. <https://doi.org/10.5334/dsj-2019-006>
- Schmiege, B., Brandt, N., Schnepf, V. J., Radosevic, L., Gretzinger, S., Selzer, M., & Hubbuch, J. (2022). Structured Data Storage for Data-Driven Process Optimisation in Bioprinting. *Applied Sciences (Switzerland)*, 12(15). <https://doi.org/10.3390/app12157728>
- Schopf, J. M., Turner, K., Doyle, D., Lake, A., Leigh, J., & Tierney, B. L. (2022). The NetSage measurement and analysis framework in practice. *Cluster Computing*, 25(4), 2967–2990. <https://doi.org/10.1007/s10586-021-03417-x>
- Schöpfel, J., Ferrant, C., André, F., Fabre, R., Schöpfel, J., André, F., & Fabre, R. (2018). *Research data management in the French National Research Center ( CNRS )*. <https://doi.org/10.1108/DTA-01-2017-0005>
- Schöpfel, J., Prost, H., & Rebouillat, V. (2017). Research Data in Current Research Information Systems. *Procedia Computer Science*, 106, 305–320. <https://doi.org/10.1016/j.procs.2017.03.030>
- Schröder, M., Staehle, S., Krüger, F., Groth, P., Nebe, J. B., & Spors, S. (2022). *Structure-based knowledge acquisition from electronic lab notebooks for research data provenance documentation*. 1–22.
- SECRETARIA - SENADO. (2009). *Leyes Desde 1992 - Vigencia Expresa Y Control De*. [http://www.secretariasenado.gov.co/senado/basedoc/ley\\_1286\\_2009.html](http://www.secretariasenado.gov.co/senado/basedoc/ley_1286_2009.html)
- Senagi, K., & Tonnang, H. E. Z. (2022). A Novel Tightly Coupled Information System for Research Data Management. *Electronics (Switzerland)*, 11(19). <https://doi.org/10.3390/electronics11193196>
- Shipman, J. P., & Tang, R. (2019). *The collaborative creation of a Research Data Management Librarian Academy ( RDMLA )*. 39, 243–247. <https://doi.org/10.3233/ISU-190050>
- Sivertsen, G. (2019). Developing current research information systems (CRIS) as data sources for studies of research. *Springer Handbooks*, 667–683. [https://doi.org/10.1007/978-3-030-02511-3\\_25/FIGURES/1](https://doi.org/10.1007/978-3-030-02511-3_25/FIGURES/1)
- Štebe, J., Dolinar, M., Bezjak, S., & Inkret, A. (2020). Implementing the rda research data policy framework in slovenian scientific journals. *Data Science Journal*, 19(1), 1–15. <https://doi.org/10.5334/dsj-2020-049>
- Stocker, M., Darroch, L., Krahl, R., Habermann, T., Devaraju, A., Schwardmann, U., D'onofrio, C., & Häggström, I. (2020). Persistent identification of instrument. *Data Science Journal*, 19(1), 1–12. <https://doi.org/10.5334/dsj-2020-018>
- Subirats-coll, I., Kolshus, K., Turbati, A., Stellato, A., Mietzsch, E., Martini, D., & Zeng, M. (2022). Original papers AGROVOC: The linked data concept hub for food and agriculture. *Computers and Electronics in Agriculture*, 196(March), 105965.

<https://doi.org/10.1016/j.compag.2020.105965>

- Suhr, M., Lehmann, C., Bauer, C. R., Bender, T., Knopp, C., Freckmann, L., Hansen, B. Ö., & Henke, C. (2020). Menoci: lightweight extensible web portal enhancing data management for biomedical research projects. *BMC Bioinformatics*, 1–21. <https://doi.org/10.1186/s12859-020-03928-1>
- Tang, R., & Hu, Z. (2019). Providing Research Data Management ( RDM ) Services in Libraries : Preparedness , Roles , Challenges , and Training for RDM Practice. *Data and Information Management*, 3(2), 84–101. <https://doi.org/10.2478/dim-2019-0009>
- Thessen, A. E., Woodburn, M., Koureas, D., Paul, D., Conlon, M., Shorthouse, D. P., & Ramdeen, S. (2019). *Proper Attribution for Curation and Maintenance of Research Collections : Metadata Recommendations of the RDA / TDWG Working Group*. 1–11.
- Töwe, M., & Barillari, C. (2020). *Who Does What ? – Research Data Management at ETH Zurich*. 1–6.
- Tsueng, G., Cano, M. A. A., Bento, J., Czech, C., Kang, M., Pache, L., Rasmussen, L. V., Savidge, T. C., Starren, J., & Wu, Q. (2023). *Developing a standardized but extendable framework to increase the findability of infectious disease datasets*. 1–13. <https://doi.org/10.1038/s41597-023-01968-9>
- UNESCO. (1998). *La Educación superior en el siglo XXI, visión y acción: informe final - UNESCO Digital Library*. 1–141. [https://unesdoc.unesco.org/ark:/48223/pf0000116345\\_spa](https://unesdoc.unesco.org/ark:/48223/pf0000116345_spa)
- Van De Sandt, S., Dallmeier-Tiessen, S., Lavasa, A., & Petras, V. (2019). The definition of reuse. *Data Science Journal*, 18(1), 1–19. <https://doi.org/10.5334/dsj-2019-022>
- Verma, L., & Kumar, N. (2018). Comparative analysis of open source digital library softwares: A case study. *DESIDOC Journal of Library and Information Technology*, 38(5), 362–369. <https://doi.org/10.14429/djlit.38.5.12425>
- Wasner, C., Barkow, I., & Odoni, F. (2018). Enhancing the research data management of computer-based educational assessments in Switzerland. *Data Science Journal*, 17, 1–6. <https://doi.org/10.5334/dsj-2018-018>
- Wiljes, C., & Cimiano, P. (2019). Teaching research data management for students. *Data Science Journal*, 18(1), 1–9. <https://doi.org/10.5334/dsj-2019-038>
- Wilkinson, M. D. (2016). *Comment: The FAIR Guiding Principles for scientific data management and stewardship*. 1–9. <https://doi.org/10.1038/sdata.2016.18>
- Xu, Z., Zhou, X., Kogut, A., & Clough, M. (2022). Effect of online research data management instruction on social science graduate students' RDM skills. *Library and Information Science Research*, 44(4), 101190. <https://doi.org/10.1016/j.lisr.2022.101190>
- Yan, C., Li, H., Pu, R., Deeprasert, J., & Jotikasthira, N. (2022). Knowledge mapping of research data in China: a bibliometric study using visual analysis. *Library Hi Tech*, 2053002. <https://doi.org/10.1108/LHT-11-2020-0285>

zenodo.org. (2023). *About Zenodo*. Zenodo.Org. <https://about.zenodo.org/>

Zervas, M., Kounoudes, A., Artemi, P., & Giannoulakis, S. (2019). Next generation Institutional Repositories: The case of the CUT Institutional Repository KTISIS. *Procedia Computer Science*, 146, 84–93. <https://doi.org/10.1016/j.procs.2019.01.083>

Zuiderwijk, A., & Spiers, H. (2019). Sharing and re-using open data: A case study of motivations in astrophysics. *International Journal of Information Management*, 49(June), 228–241. <https://doi.org/10.1016/j.ijinfomgt.2019.05.024>

Zwölf, C. M., Moreau, N., Ba, Y. A., & Dubernet, M. L. (2019). Implementing in the VAMDC the new paradigms for data citation from the research data alliance. *Data Science Journal*, 18(1), 1–13. <https://doi.org/10.5334/dsj-2019-004>



**Universidad del Cauca**  
**Facultad de Ingeniería Electrónica y Telecomunicaciones**  
**TRABAJO DE GRADO – MODALIDAD TRABAJO DE INVESTIGACIÓN**  
**ACTA DE TITULARIDAD DE DERECHOS DE PROPIEDAD INTELECTUAL PARA**  
**TRABAJOS DE GRADO, MODALIDAD INVESTIGACIÓN**

El Trabajo de Grado titulado: Repositorio de datos y sistemas de información para la gestión de la investigación: proyecto piloto en un ámbito universitario, cuyo objetivo es: Desarrollar un prototipo de repositorio de datos de investigación para la publicación de conjuntos de datos vinculado a elementos de los actuales sistemas de información de investigación (CRIS) para la gestión de grupos de investigación de la Universidad del Cauca, el cual se desarrollará por el (los) estudiante(s): Adrián Felipe Burbano Narváez y Guido Naim Imbachi Alvarado, perteneciente(s) al programa de Ingeniería de Sistemas de la Facultad de Ingeniería Electrónica y Telecomunicaciones de la Universidad del Cauca y bajo la dirección de Sandra Milena Roa Martínez, Docente adscrito(a) al Departamento de Sistemas, se realizará como requisito para optar al título de Ingenieros de Sistemas. La idea original del proyecto es de Sandra Milena Roa Martínez, identificada con C.C.63.516.527 de Bucaramanga quien la presentó al mencionado Departamento y fue aceptada como tema para el proyecto de grado en referencia.

- Financiación:

Entidades que Participan en la financiación del proyecto	
1. Porcentaje Financiación UNIVERSIDAD DEL CAUCA	100 %
2. Porcentaje Financiación entidades externas	0 %
3. Nombre de la(s) entidad(es) financiadora(s) (de ser el caso)	

Los resultados que surjan del desarrollo de este Trabajo de Grado estarán regidos por el Estatuto de Propiedad Intelectual de la Universidad del Cauca, el cual se encuentra regulado por el Acuerdo 008 de 1999, modificado por el Acuerdo Superior 004 de 2018.

Todos los participantes manifiestan que conocen y aceptan lo dispuesto en el Acuerdo Superior No. 008 de 1999 y sus respectivas modificaciones, por lo cual, los aspectos no establecidos en la presente acta serán definidos de conformidad con el Estatuto de Propiedad intelectual de la Universidad del Cauca.

En constancia de aceptación, se firma el Acta por los que en ella intervienen, a los \_\_\_\_ días del mes de enero del 2023.

\_\_\_\_\_  
Nombre: Guido Naim Imbachi Alvarado  
Estudiante  
Cédula: 1058974102

\_\_\_\_\_  
Nombre: Adrián Felipe Burbano Narváez  
Estudiante  
Cédula: 1061792924

\_\_\_\_\_  
Nombre: Sandra Milena Roa Martínez  
Director

\_\_\_\_\_  
Nombre: Alejandro Toledo Tovar  
Decano FIET

## 8. ANEXOS

### **Anexo N° 1: Síntesis y Estadísticas de Artículos**

Este anexo proporciona una visión detallada de los artículos revisados en el marco de este trabajo de grado. Cada artículo está acompañado de un resumen descriptivo y clasificaciones pertinentes. Además, se incluyen gráficas estadísticas que ofrecen una representación visual de las tendencias y distribuciones encontradas en la revisión. Haga clic en el siguiente enlace para visualizar el anexo 1: [https://docs.google.com/spreadsheets/d/1ewqiX\\_DKCQGK772KXBluXTQt2uwDv30AippV4uLyNQs/edit?usp=sharing](https://docs.google.com/spreadsheets/d/1ewqiX_DKCQGK772KXBluXTQt2uwDv30AippV4uLyNQs/edit?usp=sharing)

### **Anexo N°2: Manual de usuario**

Este anexo presenta el Manual de Usuario correspondiente al prototipo de repositorio de datos de investigación desarrollado para la Universidad del Cauca (INVESTUNICAUCA). El manual proporciona orientación detallada sobre configuración, y el uso eficiente de las funciones y características disponibles en la plataforma. Se incluyen instrucciones paso a paso, capturas de pantalla ilustrativas y consejos prácticos para garantizar una experiencia de usuario óptima.

Para acceder al manual de usuario, consulte en [https://docs.google.com/document/d/130M6BKsXOMnES1iCrS5D\\_zXWFotmTqU0Zbdfjt4yNrl/edit?usp=sharing](https://docs.google.com/document/d/130M6BKsXOMnES1iCrS5D_zXWFotmTqU0Zbdfjt4yNrl/edit?usp=sharing) . Este recurso es esencial para cualquier persona que desee comprender y aprovechar al máximo las capacidades del prototipo.