



Esta obra está bajo una [Licencia Creative Commons Atribución - 4.0 Internacional \(CC BY 4.0\)](https://creativecommons.org/licenses/by/4.0/)

Vea una copia de esta licencia en <https://creativecommons.org/licenses/by/4.0/deed.es>





FACULTAD DE INGENIERÍA DE SISTEMAS E INFORMÁTICA
ESCUELA PROFESIONAL DE INGENIERÍA DE SISTEMAS E INFORMÁTICA

Tesis

Agregación de repositorios institucionales para la generación de información del desempeño científico de universidades peruanas

Para optar el título profesional de Ingeniero de Sistemas e Informática

Autor:

Edy Luis Huancaruna Golac
<https://orcid.org/0000-0002-7492-0649>

Asesor:

Ing. Dr. Miguel Ángel Valles Coral
<https://orcid.org/0000-0002-8806-2892>

Tarapoto, Perú

2023



FACULTAD DE INGENIERÍA DE SISTEMAS E INFORMÁTICA
ESCUELA PROFESIONAL DE INGENIERÍA DE SISTEMAS E INFORMÁTICA

Tesis

Agregación de repositorios institucionales para la generación de información del desempeño científico de universidades peruanas

Para optar el título profesional de Ingeniero de Sistemas e Informática

Autor:

Edy Luis Huancaruna Golac

Sustentado y aprobado el 27 de octubre del 2023, por los jurados:

Presidente de Jurado

Ing. M. Sc. Jorge Damián Valverde
Iparraguirre

Secretario de Jurado

Ing. John Clark Santa Maria
Pinedo

Vocal de Jurado

Ing. Mtro. Cristian Werner García
Estrella

Asesor

Ing. Dr. Miguel Ángel Valles
Coral

Tarapoto, Perú

2023



Universidad Nacional de San Martín
Facultad de Ingeniería de Sistema e Informática
Jr. Vía Universitaria S/Nº - Ciudad Universitaria - Morales
Teléf. (042) 525688 - 524074 - Anexo 109



ACTA DE SUSTENTACIÓN PARA OPTAR EL TÍTULO DE INGENIERO DE SISTEMAS E INFORMÁTICA

En los ambientes del Aula Magna de la Facultad de Ingeniería de Sistemas e Informática de la Universidad Nacional de San Martín, a las 08:00 horas del día viernes 27 de octubre del año 2023, se reunieron los **miembros del Jurado Calificador**, integrado por:

Presidente : **ING. DR. JORGE DAMIAN VALVERDE IPARRAGUIRRE**
Secretario : **ING. JOHN CLARK SANTA MARIA PINEDO**
Vocal : **ING. Mtro. CRISTIAN WERNER GARCÍA ESTRELLA**

Para evaluar la Tesis: "AGREGACIÓN DE REPOSITORIOS INSTITUCIONALES PARA LA GENERACIÓN DE INFORMACIÓN DEL DESEMPEÑO CIENTÍFICO DE UNIVERSIDADES PERUANAS" presentada por el Bachiller **EDY LUIS HUANCARUNA GOLAC**, participando en calidad de asesor el Ing. Dr. Miguel Ángel Valles Coral.

Los señores miembros del Jurado, después de haber atendido la sustentación y evaluada las respuestas a las preguntas formuladas y terminada la réplica; luego de debatir entre sí, reservada y libremente lo declaran aprobado por unanimidad, con el calificativo de excelente, equivalente a diecinueve (19), en fe de lo cual firmamos la presente acta, siendo las 9:00 horas del mismo día, con lo que se dio por terminado el Acto de Sustentación.


.....
**ING. DR. JORGE DAMIAN VALVERDE
IPARRAGUIRRE**
Presidente


.....
**ING. JOHN CLARK SANTA MARIA
PINEDO**
Secretario


.....
**ING. Mtro. CRISTIAN WERNER GARCÍA
ESTRELLA**
Vocal

Declaratoria de autenticidad





Edy Luis Huancaruna Golac, con DNI N° 71472696, egresado de la Escuela Profesional de Ingeniería de Sistemas e Informática Facultad de Ingeniería de Sistemas e Informática de la Universidad Nacional de San Martín, autor de la tesis titulada: Agregación de repositorios institucionales para la generación de información del desempeño científico de universidades peruanas.

Declaramos bajo juramento que:

1. La tesis presentada es de mi autoría.
2. La redacción fue realizada respetando las citas y referencia de las fuentes bibliográficas consultadas, siguiendo las normas APA actuales
3. Toda información que contiene la tesis no ha sido plagiada;
4. Los datos presentados en los resultados son reales, no han sido alterados ni copiados, por tanto, la información de esta investigación debe considerarse como aporte a la realidad investigada.

Por lo antes mencionado, asumo bajo responsabilidad las consecuencias que deriven de mi accionar, sometiéndome a las leyes de nuestro país y normas vigentes de la Universidad Nacional de San Martín.

Tarapoto, 27 de octubre de 2023.

 	 
Edy Luis Huancaruna Golac 71472696 Autor	Ing. Dr. Miguel Ángel Valles Coral 40810431 Visto bueno Asesor

Ficha de identificación

<p>Título del proyecto Agregación de repositorios institucionales para la generación de información del desempeño científico de universidades peruanas</p>	<p>Área de investigación: Ingeniería y Tecnología Línea de investigación: Ciencias de la computación Sublínea de investigación: Inteligencia artificial y recuperación de la información Grupo de investigación: GIIA RESOLUCIÓN N° 134–2021–UNSM/FISI/CFT Tipo de investigación: Básica <input type="checkbox"/>, Aplicada <input checked="" type="checkbox"/>, Desarrollo experimental <input type="checkbox"/></p>
<p>Autor: Edy Luis Huancaruna Golac</p>	<p>Facultad de Ingeniería de Sistemas e Informática Escuela Profesional de Ingeniería de Sistemas e Informática https://orcid.org/0000-0002-7492-0649</p>
<p>Asesor: Ing. Dr. Miguel Ángel Valles Coral</p>	<p>Dependencia local de soporte: Facultad de Ingeniería de Sistemas e Informática Escuela Profesional de Ingeniería de Sistemas e Informática Unidad o Laboratorio Ingeniería de Sistemas e Informática https://orcid.org/0000-0002-8806-2892</p>

Dedicatoria

Dedico este trabajo a mi madre Magally Golac Mori, cuyo amor, sacrificio y confianza inquebrantable en mí me han impulsado a alcanzar mis metas académicas. Su apoyo incondicional ha sido mi mayor motivación.

Dedico este trabajo a mi hermana Leydi Díaz Golac, quien ha sido mi fuente constante de inspiración y ha compartido mis triunfos y desafíos a lo largo de esta carrera académica. Su apoyo y aliento significan el mundo para mí.

Esta tesis está dedicada a mis bisabuelos, cuyo ejemplo de perseverancia y determinación siempre ha sido una fuente de inspiración para mí. Aunque uno de ellos no está físicamente presente, sé que estaría orgullosa de este logro.

Agradecimientos

Quiero expresar mi profunda gratitud a mi asesor de tesis, Ing. Dr. Miguel Ángel Valles Coral por su orientación experta, apoyo constante y valiosos comentarios a lo largo de este proceso.

Agradezco sinceramente a mi familia por su amor incondicional, paciencia y apoyo inquebrantable. Sin ellos, este logro no habría sido posible.

Me gustaría agradecer a mis amigos y compañeros de clase por su aliento, motivación y por compartir sus conocimientos conmigo durante nuestra trayectoria académica juntos.

Expreso mi agradecimiento a todos los profesores y personal administrativo de la universidad por su dedicación y compromiso en brindar una educación de calidad.

Índice general

Ficha de identificación	6
Dedicatoria	7
Agradecimientos.....	8
Índice general.....	9
Índice de tablas	11
Índice de figuras	12
RESUMEN	13
ABSTRACT	14
CAPÍTULO I INTRODUCCIÓN A LA INVESTIGACIÓN.....	15
CAPÍTULO II MARCO TEÓRICO	18
2.1. Antecedentes de la investigación	18
2.2. Fundamentos teóricos.....	20
2.3. Definición de términos básicos.....	23
CAPÍTULO III MATERIALES Y MÉTODOS.....	25
3.1. Ámbito y condiciones de la investigación.....	25
3.1.1 Contexto de la investigación.....	25
3.1.2 Periodo de ejecución.....	25
3.1.3 Autorizaciones y permisos	25
3.1.4 Control ambiental y protocolos de bioseguridad.....	25
3.1.5 Aplicación de principios éticos internacionales.....	25
3.2. Sistema de variables	26
3.2.1 Variables principales	26
3.3 Procedimientos de la investigación.....	26
3.3.1 Objetivo específico 1: Extraer los metadatos de los RI de las universidades peruanas mediante protocolos de interoperabilidad	27
3.3.2 Objetivo específico 2: Agregar y Procesar los datos en un sistema de información general basado en dashboard para la visualización de la información de la producción científica de las universidades.....	28

3.3.3	Objetivo específico 3: Evaluar el desempeño científico de los RI y plantear mejoras a partir de la información generada.....	28
CAPÍTULO IV RESULTADOS Y DISCUSIÓN		29
4.1	Resultado específico 1	29
4.2	Resultado específico 2	32
4.3	Resultado específico 3	40
CONCLUSIONES.....		45
RECOMENDACIONES.....		46
REFERENCIAS BIBLIOGRÁFICAS.....		47

Índice de tablas

Tabla 1 Descripción de variables	26
Tabla 2 Cantidad de repositorios habilitados al momento de la validación	29
Tabla 3 Cantidad de Repositorios que cuentan con SSL activado	29
Tabla 4 Cantidad de Repositorios que cuentan con SSL activado	30
Tabla 5 Cantidad de elementos extraídos	30
Tabla 6 Total de elementos por tipo de etiqueta	33
Tabla 7 Top 10 Universidades.....	34
Tabla 8 Número de elementos por tipo de publicación.....	35
Tabla 9 Top 10 Asesores agrupación sin filtro	36
Tabla 10 Top 10 Asesores agrupación con filtro	37
Tabla 11 Comparación de Número de publicaciones por Tipo de Acceso.	37
Tabla 12 Top 10 Tesistas sin filtros.	38
Tabla 13 Top 10 Tesistas con filtros.....	38
Tabla 14 Cantidad de elementos por Región.	43

Índice de figuras

Figura 1 Diagrama de flujo de la ejecución del proyecto	27
Figura 2 Estructura de los metadatos.....	29
Figura 3 Dashboard parte 1 sin filtros: tipo de dato y tipo de acceso.	33
Figura 4 Dashboard parte 1 con filtros: tipo de dato y tipo de acceso.	33
Figura 5 Dashboard parte 2 sin filtro: Top 10 universidades y Tipo Publicación.	34
Figura 6 Dashboard parte 2 con filtro: Top 10 universidades y Tipo Publicación.	34
Figura 7 Dashboard parte 3 sin filtros: tipo de dato y tipo de acceso.	36
Figura 8 Dashboard parte 3 con filtros: tipo de dato y tipo de acceso.	36
Figura 9 Dashboard parte 4 sin filtros: Top 10 Tesistas.	37
Figura 10 Dashboard parte 4 con filtros: Top 10 Tesistas.	38
Figura 11 Dashboard parte 5 sin filtros: tipo de gestión.	39
Figura 12 Dashboard parte 5 con filtros: tipo de gestión.	39
Figura 13 Macrorregión Norte	40
Figura 14 Macrorregión Centro	40
Figura 15 Macrorregión Sur.....	41
Figura 16 Macrorregión Oriente	41
Figura 17 Región Costa.....	42
Figura 18 Región Sierra	42
Figura 19 Región Selva	43

RESUMEN

Agregación de repositorios institucionales para la generación de información del desempeño científico de universidades peruanas.

Al pasar los años, la cantidad de material académico continua en aumento, generando una gran cantidad de datos los cuales se almacena directamente en los repositorios institucionales, esta gran cantidad de datos no es aprovechada correctamente por los gestores de los RI, pues los métodos de análisis no entregan una información de alto valor, limitando de esta manera la toma de decisiones. El objetivo es generar información del desempeño científico de las universidades peruanas mediante la agregación de sus repositorios institucionales, implementando un sistema de información gerencial basado en dashboard para la generación de información, el cual permitió la evaluación del desempeño científico de las universidades. Esta investigación abarca las universidades licenciadas del Perú registradas en SUNEDU. La investigación es de tipo no experimental, descriptivo; la cual tiene como procedimiento principal la extracción de los metadatos de los RI usando protocolos de interoperabilidad mediante web scraping; Los metadatos obtenidos fueron procesados y visualizados en un dashboard. Los resultados obtenidos mostraron que la mayor cantidad de universidades que mantienen su RI habilitado son de un tipo gestión privada, además que una gran cantidad de universidades no mantienen una correcta gestión en sus metadatos, de tal forma se determinó que, en la región norte y centro, así como la costa del país se encuentra la mayor cantidad de material académico, lo cual refleja un mejor acceso a recursos científicos y académicos. Se concluyo que la gran cantidad de datos dificulta la visualización de los metadatos, por lo cual se recomienda que en futuras investigaciones se realice la visualización de datos enfocándose en soluciones big data.

Palabras clave: Repositorio Institucional, OAI, Dashboard, análisis de datos, visualización de información.

ABSTRACT

Aggregation of institutional repositories for the generation of information on the scientific performance of Peruvian universities.

Over the years, the amount of academic material continues to increase, generating a large amount of data which is stored directly in institutional repositories; this large amount of data is not properly exploited by IR managers, because the analysis methods do not deliver high-value information, thus limiting decision making. The objective is to generate information on the scientific performance of Peruvian universities by aggregating their institutional repositories, implementing a dashboard-based management information system for the generation of information, which allowed the evaluation of the universities' scientific performance. This research covers Peruvian licensed universities registered in SUNEDU. The research is non-experimental, descriptive; its main procedure is the extraction of IR metadata using interoperability protocols through web scraping. The metadata obtained were processed and visualized in a dashboard. The results obtained showed that most of the universities that maintain their IR enabled belong to a private management type, and that a large number of universities do not maintain a correct management of their metadata. In this way, it was determined that the northern and central regions, as well as the coast of the country, have the largest amount of academic material, which reflects a better access to scientific and academic resources. It was concluded that the large amount of data hinders the visualization of metadata, so it is recommended that in future research, data visualization should focus on big data solutions.

Keywords: Institutional Repository, OAI, Dashboard, data analysis, information visualization.



CAPÍTULO I INTRODUCCIÓN A LA INVESTIGACIÓN

Los Repositorios Institucionales (RI) se identifican como una plataforma que brinda acceso abierto a productos académicos (Bamigbola y Adetimirin, 2020); dando soporte al aprendizaje y la investigación de estudiantes y docentes universitarios (Gunderman, Scherer y Behrman, 2021). Así mismo, ofrecen una solución a la gestión del conocimiento con la recolección de metadatos, los cuales facilitan la interoperabilidad (Chisita y Chiparausha, 2021). Si bien, dicha recolección permite monitorear los datos, puede ser algo complicado acceder a sus estadísticas, debido a que no cuenta con las funcionalidades necesarias (Vardell, 2020).

En su mayoría las universidades tienen implementado un RI para el almacenamiento y la diseminación de productos académicos e intelectuales producidos por su comunidad. Las peruanas no son la excepción a este caso, pues hasta junio de 2020 más del 80% de las 92 universidades licenciadas por SUNEDU cuentan con un repositorio institucional (Blanco-Olea, 2021a).

En Perú las universidades licenciadas disponen de un RI para la gestión de su material académico; el cual es almacenado en conjunto con sus metadatos y cumplen un papel importante, pues permite mejor administración en la recuperación de datos (White y Radio, 2020). Sin embargo, los RI de estas universidades no cuentan con funciones que permitan visualizar la agregación de los metadatos desde una perspectiva gráfica y que permita conocer indicadores del desempeño de las unidades y/o dependencias funcionales de la institución a la que el repositorio pertenece.

En un enfoque centrado a la administración de los metadatos en los RI, el administrador es el que tiene la tarea de identificar y gestionar los datos en los productos académicos, pues los investigadores, editores y otras partes interesadas, buscan publicar investigaciones para de esta manera recibir reconocimiento. Por este motivo es crucial que el administrador sepa identificar y monitorear los datos de investigadores y sus actividades académicas (Mering, 2019). Ya que al generarse inconformidad con el RI los investigadores podrían preferir publicar sus investigaciones en otros medios o portales de su preferencia (Price y Murtagh, 2020).

Si bien la gestión de los RI se trata de un proceso constante, el cual se basa en la difusión de contenido y la administración de material académico, un gran volumen de datos requiere una constante supervisión y la correcta toma de decisiones sobre la gestión de los datos, elementos, campo de los metadatos y el tratamiento de los mismos

(Shelley, 2020). Además, se identificó que los administradores de los RI no cuentan con un amplio conocimiento de los procesos y mantenimiento de las funciones del RI (Masinde y Sanya, 2022).

Se identifica que, a pesar de la buena implementación de tecnología de información para los RI, el cual recompila, preserva y brinda acceso abierto a una gran cantidad de datos de investigaciones y productos académicos, estos no cuentan con funcionalidades que permitan identificar los datos y procesos del material académico que se publica. Al aplicar servicios de tecnología en los proyectos e investigaciones de autoría interna, se aprovechan mejor los mecanismos y los casos de uso en los que se emplean la tecnología (Gunderman, Scherer y Behrman, 2021).

La inexistencia de herramientas que permitan visualizar gran cantidad de datos en las universidades peruanas, afecta negativamente a la interacción que puedan tener los investigadores con los RI. La plataforma web de cada una de ellas, presenta un apartado donde se puede visualizar información estadística de muy poco valor, el cual no contribuye a sus representantes a tener un buen análisis.

Distintas organizaciones han integrado el análisis de datos al proceso administrativo (Schoen et al., 2019). El cual contribuye en gran medida a la toma de decisiones (Wu, Dinçer y Yüksel, 2022), centrándose principalmente en obtener la mayor cantidad de beneficios a partir de la información generada. Sin embargo, un gran volumen de información trae consigo, un proceso muy complicado de análisis (Vázquez-Ingelmo, García-Peñalvo y Therón, 2021). Como menciona (Susnjak, Ramaswami y Mathrani, 2022) la toma de decisiones con la ayuda especialmente de una capa de información, permite a los usuarios tener una perspectiva más amplia al momento de decidir una estrategia.

No tener en cuenta el conocimiento de las personas que administran los RI y/o autoridades de la universidad, tiende a ser un problema muy común. Ya que se determina, que su percepción es sumamente variable, pues existe un grupo que admite estar empapado de conocimiento o quienes mencionan estar informados de la existencia de dicha información, pero no sobre el accionar que plantean las universidades (Andrés et al., 2020).

Bajo este panorama, se pretendió implementar un panel de análisis de datos como herramienta principal para la toma de decisiones, con la visualización de los datos almacenados en los RI de las Universidades del Perú, empleando métodos de visualización tipo dashboard, de manera que los encargados del repositorio y cualquier individuo vinculado a la investigación pueda plantear estrategias. De tal forma dentro de

este marco, se planteó la hipótesis: La agregación de los repositorios institucionales permitirá generar información del desempeño científico de las universidades peruanas.

Del mismo modo, se propuso como objetivo general generar información del desempeño científico de las universidades peruanas mediante la agregación de sus repositorios institucionales. A continuación, los objetivos específicos fueron: 1) Extraer los metadatos de los RI de las universidades peruanas mediante protocolos de interoperabilidad; 2) Agregar y Procesar los datos en un sistema de información general basado en dashboard para la visualización de la información de la producción científica de las universidades; 3) Evaluar el desempeño científicos de los RI y plantear mejoras a partir de la información generada.

CAPÍTULO II MARCO TEÓRICO

2.1. Antecedentes de la investigación

En su investigación (Lock, Bednarz y Pettit, 2020) identificaron, que los datos sin procesar son un problema evidente en las ciudades de gran tamaño, los cuales son generados por una serie de sensores y dispositivos en la ciudad. Por esta razón los autores propusieron la implementación de un dashboard el cual tiene como objetivo brindar servicios que permitan a los clientes y administradores del transporte estar constantemente informados. Para lograr esto se implementó una base de datos sobre todos los datos sin procesar recaudados en el transcurso de un año, dichos datos eran filtrados y limpiados mediante scripts en Python, para ser visualizados y comparados con datos actuales mediante un dashboard. Los resultados de la investigación concluyen con recomendaciones sobre la importancia de este enfoque en casos similares basados en la planificación de transporte, gestión de rendimiento, optimización y toma de decisiones.

En su estudio (Dobraja et al., 2020) menciona que, a raíz de la gran cantidad de datos de movimiento, es necesario no solo tener conocimientos en extracción y almacenamiento de estos datos, sino que también en cómo representarlos visualmente. Se identifica al dashboard como una solución apropiada para la petición de una gran cantidad de datos y diferentes variables, pero con una limitante muy relevante, la flexibilidad del diseño. Por esta razón el estudio tiene como objetivo implementar la adaptabilidad en los dashboards para garantizar que los usuarios obtengan información relevante y adecuada a su interés, con varios niveles de detalle. El estudio basó su enfoque en el diseño centrado en el usuario, con el fin de que los usuarios que interactúan con el dashboard no se sientan abrumados con la gran cantidad de datos que se muestran. Además, se concluyó que los tableros adaptables permiten que los usuarios tengan mayor participación en el proceso de obtención de información.

En su investigación (Turenne, 2023) menciona como las redes sociales son cada vez más dominantes en la sociedad de tal manera que el contenido de YouTube, puede ser un recurso muy útil en la ciencia computacional para comprender las preguntas claves sobre la sociedad. Por lo cual su investigación realizó una extracción de datos de los contenidos de YouTube usando web scraping para crear un conjunto de datos de 644575 transcripciones de video. Para esto se realizó la extracción automática de características lingüísticas para capturar una representación de cada video usando título, descripción y transcripción, los cuales son entendidos como metadatos descargados. Luego se realizó una limpieza de metadatos mediante la agrupación

automática con representación lingüística con el fin de identificar videos no coincidentes y palabras clave. Usando estas palabras clave se crean filtros para excluir videos y finalmente se obtuvo un conjunto de datos la cual se redujo un 95%. Los resultados de la investigación otorgan un dataset sobre metadatos de videos de YouTube el cual con ayuda de análisis de datos se puede explotar para estudios de sociología digital y democracia electrónica.

En su investigación (Scott, 2022) menciona que mientras buscaba comprender el impacto en el legado de Alma Mahler compositora y editora musical austríaca, identifico la necesidad de recopilar la mayor cantidad de datos posible sobre la interpretación de sus canciones. Esta necesidad origino la evaluación de metadatos asociados con las grabaciones de las canciones de dicha autora, los cuales se encontraban en WorldCat y Youtube. La investigación relata un enfoque en el método de web scraping de datos de YouTube destacando algunas promesas y limitaciones asociadas con el web scraping. Realizo extracción en ambas plataformas, con la diferencia que con los datos de YouTube se empleó automatización, por la gran cantidad de datos que esta plataforma posee. Usando DataMiner una solución que permite extraer datos de sitios web, se extrajeron los metadatos necesarios usando una configuración de puntos de interés, en los cuales se recopilaban la URL, el título, la cantidad de visitas, la fecha de publicación, el canal y la descripción. Luego se analizaron y compararon los datos obtenidos para determinar que el análisis de estos datos complementa los métodos musicológicos más tradicionales y respalda una comprensión más holística de la interpretación musical contemporánea.

Así mismo, en Asia (Ghosh y Roy, 2021) examinan y analizan 88 repositorios digitales de acceso abierto, los cuales están centrados en la ciencia agrícola, además de determinar cuál es el impacto que estos presentan en la web. Pues gracias a la iniciativa de acceso abierto a la información, los repositorios propagan información a usuarios en varias instituciones, por lo cual es preciso una evaluación de la influencia que tienen en la web. Tras realizado el análisis de los datos que fueron recopilados mediante el método de encuesta y observación, concluyeron que a pesar de la cantidad de páginas web que los repositorios contienen, el cual resalta su impacto, los efectos son cambiantes debido a problemas en la estructura del sitio web de los repositorios y la baja indexación en los directorios web. Por lo cual recomiendan, un mayor seguimiento por parte de los administradores para así regular el aumento de la presencia del repositorio en la web.

De acuerdo a (Bréant et al., 2020) en su investigación identifica como los hospitales poseen entornos muy complejos y variados lo que genera una gran dificultad en la medición del desempeño de sus actividades, por tal motivo menciona la importancia de la implementación de los dashboards como una herramienta de monitoreo y análisis para mejorar la calidad en la atención, así como los procesos administrativos y clínicos; el objetivo de su investigación es proporcionar información que sea fácil de comprender por personal médico y administrativo como también la contribución en la toma de decisiones para el seguimiento estratégico y operativo. Además, se determinó que el uso constante de dicha herramienta puede contribuir con la detección de situaciones críticas con una mayor eficacia.

En su investigación (Kenigsberg et al., 2022) implementó un dashboard interno el cual tenía el objetivo de facilitar la visualización y la revisión de una gran cantidad de datos sobre la vigilancia de la seguridad de la vacuna COVID-19, en los cuales analizo datos como la población, niveles de aceptación, análisis de secuelas y otros datos de interés, los cuales permitieron informar la situación actual de la cobertura de vacunación, para tener un correcto análisis y contribuir con la toma de decisiones.

2.2. Fundamentos teóricos

Open Access

Los movimientos de Open Access (acceso abierto) realizaron cambios en la comunicación académica, la cual se centra en proporcionar acceso gratuito a la literatura académica a través de internet en los repositorios institucionales de las distintas universidades y entidades gubernamentales en el mundo (Ghosh y Roy, 2021).

La Iniciativa de Acceso Abierto de Budapest menciona que el material categorizado en el acceso abierto, otorga la libertad de ser usado sin restricciones dándole crédito al autor original. Además, se indica que no tienen que existir algún problema con la copia, descarga y remezcla de información (Sainath Dandawate y Dhanmjaya, 2021).

Repositorio Institucional

Los RI se originaron como una plataforma para brindar acceso abierto a trabajos académicos, las instituciones especialmente universidades disponen de un RI para almacenar, administrar y difundir la producción intelectual de su comunidad. Este a su vez cuenta con cuatro características: definido institucionalmente, contenido académico, acumulativa y perpetua y por último interoperabilidad abierta. Desde el punto de vista institucional, los beneficios del uso de RI se consideraron como la entrega de un lugar centralizado para el almacenamiento todo tipo de producción institucional.

Adicionalmente el RI es una forma de aumentar el alcance, la disponibilidad, accesibilidad y descubrimiento del contenido académico de la institución sin ningún costo adicional (Bamigbola y Adetimirin, 2020).

Es una plataforma que se inició para poder permitir el acceso abierto a material académico en las instituciones y tiene como propósito de almacenar, archivar, difundir y permitir la administración de la producción académica de miembros de la comunidad académica (Bamigbola y Adetimirin, 2020). También se determina como un archivo digital lleno de trabajos académicos como tesis, investigaciones, actas el cual se considera como el capital intelectual que posee las universidades (Chisita y Chiparausha, 2021).

Tipos de Repositorios

Repositorios temáticos: preserva y almacena información de disciplinas o áreas temáticas a las cuales se tiene libre acceso, además tienen la particularidad de ser creados y gestionados por instituciones académicas orientadas a la investigación, o por organizaciones gubernamentales (Sanchez y Mariño, 2021).

Repositorios de datos: pueden alojar y preservar datos científicos los cuales son resultado de investigaciones, estos pueden ser de carácter independiente o pueden estar vinculados a un RI (Sanchez y Mariño, 2021).

Repositorios huérfanos: se caracterizan por estar establecidos a nivel nacional, brindan la posibilidad de publicar investigaciones a autores los cuales no tienen una relación directa con otros repositorios (Sanchez y Mariño, 2021).

Metadatos

Se denominan como la información que describe los datos de las investigaciones (Wang et al., 2021). Los datos generados por informes científicos son tomados como una base que sustenta todos los progresos en la ciencia, pero dichos datos sin una descripción carecen de valor, por lo cual se define a los metadatos como una descripción de los datos científicos (San Gil et al., 2010).

Metadatos Dublin Core

Es un esquema de metadatos el cual ha sido aprobado como un estándar por el Comité Europeo de Normalización (CEN), tiene como objetivo mejorar el proceso para encontrar recursos en internet (Greenberg, 2009).

Dashboard

Los dashboards utilizan técnicas de visualización de datos, los cuales proporcionan información de valor a personas, servicios u organizaciones, además, existe la creencia que el uso de estas técnicas de visualización mejora la comprensión de la información y sobre todo reduce la carga cognitiva (Randell et al., 2022).

Se definen como herramientas que permiten agrupar una cantidad de datos los cuales podrán ser compartidos mediante una interfaz gráfica, la cual mostrara información relevante para la organización (Córdova Viera, Martínez Borrego y Córdova Viera, 2021).

Los tableros brindan un acceso relévate a información oportuna la cual permite tener una buena gestión en la toma de decisiones, mejorando así la calidad de la institución a la cual pertenece (Bréant et al., 2020).

Visualización de datos

La visualización de datos permite dar una forma concisa a datos complejos, los cuales pueden ser visualizados a través de patrones o uniones que no pueden ser entendidas de una manera más simple cuando se visualizan los datos sin ningún procesamiento previo (Conejero et al., 2021).

Marco de visualización

Representa una serie de métodos de visualización con los cuales se puede comunicar diversas dimensiones, las cuales están relacionados a una gran cantidad de datos (Lock, Bednarz y Pettit, 2020).

A. Mapa de calor

Se definen como una representación gráfica de datos, los cuales usan un sistema de color, el cual permite la identificación de las tendencias de manera más eficiente y ayuda a dirigir a los usuarios hacia áreas con datos de alto valor (Lock, Bednarz y Pettit, 2020).

B. Nube de puntos

Es una técnica de visualización de datos el cual permite mostrar patrones de datos espaciotemporales (Lock, Bednarz y Pettit, 2020).

Diseño de dashboard

En su investigación (Ávila-Chaves, 2021) pretenden centrar el enfoque en la toma de decisiones usando información generada por el dashboard, para esto se identifican los

indicadores de desempeño los cuales, no deben perder la perspectiva de un dashboard, pues los datos en relación a los indicadores tienen que estar pensados con el fin de liderar en la toma de decisiones. Además, es importante determinar que el tablero no es una simple muestra de datos, esta tiene que estar directamente relacionado con la administración de los recursos, dicho tablero tiene que tener la particularidad de entregar información de un vistazo y simultáneamente tiene que ser entendible para cualquier usuario.

2.3. Definición de términos básicos.

ALICIA

ALICIA es el Repositorio Nacional Digital de Ciencia, Tecnología e Innovación, el cual pertenece a CONCYTEC, este repositorio nacional se encarga de recolectar y brindar a todos los usuarios acceso abierto a todo el material científico, tecnológico y de innovación, el cual es generado por miembros de la comunidad pública o del estado peruano (Blanco-Olea, 2021b).

DSpace

DSpace es una aplicación informática, la cual fue diseñada para instituciones y organizaciones como una aplicación de código abierto que permita ser ejecutada con una cantidad baja de recursos (Baudoin y Branschofsky, 2008).

Interoperabilidad

La interoperabilidad en los repositorios institucionales se determina como un componente crítico pues genera compatibilidad con otros repositorios, lo cual trae consigo un respaldo en los sistemas de comunicación académica (Wong, 2019).

Se define como la capacidad en la que múltiples sistemas, los cuales usando diferentes plataformas y estructura de datos pueden fácilmente intercambiar y compartir datos, los cuales pueden ser usados sin ningún inconveniente (San Gil et al., 2010).

OAI-PMH

Se identifica como un protocolo de Iniciativa de Archivos Abiertos para la Recolección de Metadatos (Price y Murtagh, 2020). Este protocolo permite que los repositorios puedan conectarse facilitando la interoperabilidad, además se mejoró notablemente la visibilidad que tenían los investigadores (Chisita y Chiparausha, 2021).

RECOLECTA

Se denomina como un portal que reúne a todos los repositorios españoles, el cual se encarga de brindar servicios a los administradores de repositorios, investigadores y a terceros los cuales estén relacionados con la elaboración de políticas (Blanco-Olea, 2021).

RENATI

Se define como un recopilador de investigaciones académicas las cuales son usadas para optar por un grado académico o un título profesional es cual es entregado por las universidades abaladas por SUNEDU (Blanco-Olea, 2021).

CAPÍTULO III MATERIALES Y MÉTODOS

3.1. Ámbito y condiciones de la investigación

3.1.1 Contexto de la investigación

La investigación se llevó a cabo en la Universidad Nacional De San Martín ubicado en el distrito de morales, provincia y departamento de San Martín.

3.1.2 Periodo de ejecución

La ejecución del proyecto se llevó a cabo durante los meses de noviembre 2022 a mayo 2023.

3.1.3 Autorizaciones y permisos

Dado que la extracción de metadatos se realizó mediante protocolos OAI disponibles en los repositorios institucionales, el cual permite acceso abierto de los datos a cualquier individuo.

3.1.4 Control ambiental y protocolos de bioseguridad

Contemplando la naturaleza de la investigación la cual se basa en la extracción de metadatos no fue necesario el uso de bioseguridad o se atentó contra el medio ambiente.

3.1.5 Aplicación de principios éticos internacionales

Según los resultados obtenidos de la investigación no atentan contra ningún individuo o grupo de personas en específico. Se pretende que la información obtenida se de mucha utilidad para otros investigadores y el grupo académico encargado de los RI de cada universidad, de manera que se pueda mejorar dicha gestión.

3.2. Sistema de variables

3.2.1 Variables principales

Tabla 1

Descripción de variables

Objetivo específico: Realizar la agregación de los repositorios institucionales			
Variable abstracta	Variable concreta	Medio de registro	Unidad de medida
Agregación de los repositorios institucionales	Visibilidad	Índices estadísticos	Porcentaje
	Disponibilidad	Índices estadísticos	Porcentaje
	Análisis	Índices estadísticos	Porcentaje

3.3 Procedimientos de la investigación

a) Tipo y nivel de la investigación

Esta investigación es de tipo aplicado, pues se aplican conocimientos de análisis y visualización en herramientas de interfaces gráficas, para mejorar la toma de decisiones.

El nivel de la investigación es descriptivo, pues se llevó a cabo una recopilación de metadatos de las universidades que presentan un RI, de las cuales se llevó a cabo la observación y análisis de datos para posteriormente realizar una herramienta que permita ser una base para futuros proyectos, los cuales se relacionen con la investigación.

b) Población y muestra

En esta investigación no se realizará muestreo, ya que la población consta de un grupo de 95 universidades licenciadas con repositorios institucionales de los cuales usan DSpace.

c) Diseño analítico, muestral y experimental

La investigación tiene un diseño no experimental, descriptivo.

Para llevar a cabo el proyecto fue necesario extraer los metadatos disponibles en los RI, usando un algoritmo de web scraping mediante protocolos de OAI, comprobando la estructura de los mismos.

La extracción de los datos se realizó el 20 de mayo del 2023.

Se agregó y procesó los datos obtenidos en la herramienta visual Power BI. Una vez presentado el dashboard se realizaron una serie de preguntas para evaluar el desempeño científico de las universidades con la información que se presentó al ver los datos gráficamente.

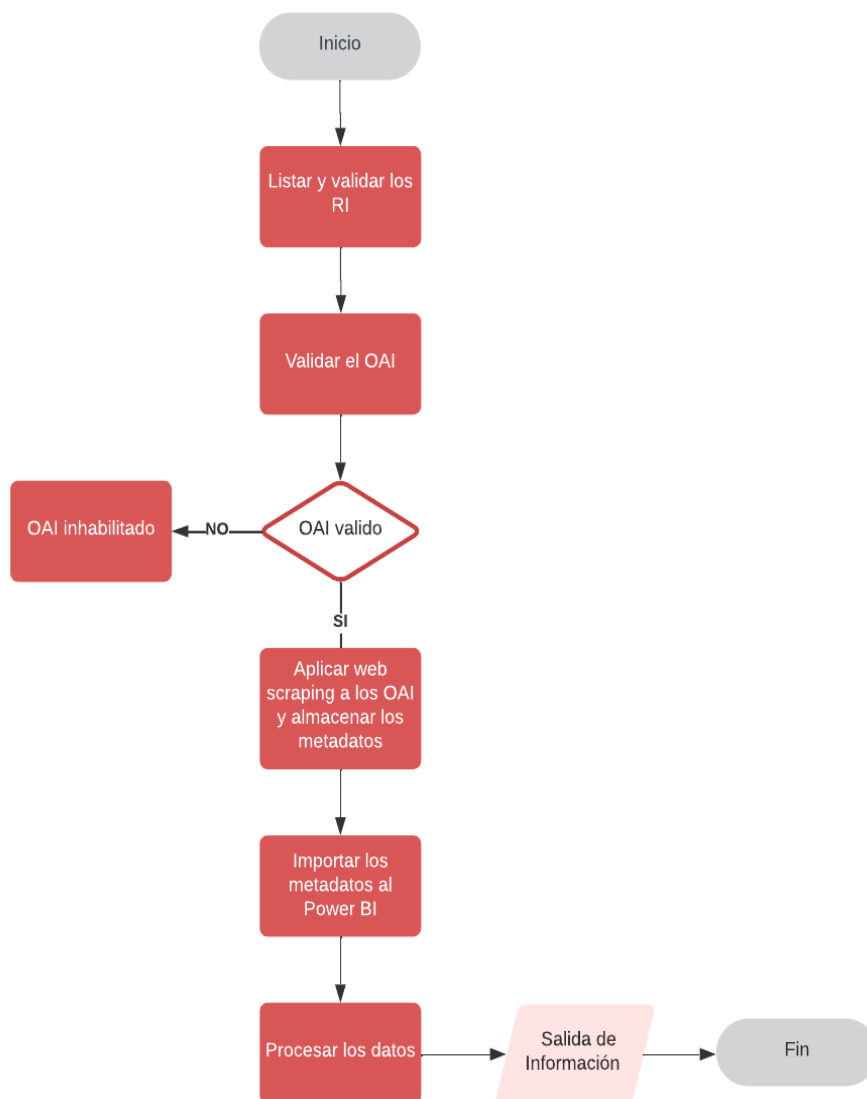


Figura 1.

Diagrama de flujo de la ejecución del proyecto

3.3.1 Objetivo específico 1: Extraer los metadatos de los RI de las universidades peruanas mediante protocolos de interoperabilidad

- Validación de los repositorios institucionales: Para realizar esta actividad ingresamos a la lista de universidades licenciadas que otorga SUNEDU, la cual nos permitió listar 95 instituciones a las cuales se validó si disponían de un RI funcional. Esta lista está disponible en: <https://docs.google.com/spreadsheets/d/1JlfxhuusJ7Hss5VflSJkYZ-Is0KCJH59/edit?usp=sharing&ouid=106909389732919702792&rtfop=true&sd=true>.
- Determinación de los metadatos requeridos: Los metadatos son la pieza fundamental en el proyecto, por tal motivo se realizó un análisis en la estructura de

los metadatos para identificar que metadatos son relevantes para el objetivo del proyecto.

- Extracción de los metadatos usando web scriping: Para realizar esta actividad codificamos el algoritmo de web scraping usando Python el cual nos permitió automatizar el proceso de extracción.

3.3.2 Objetivo específico 2: Agregar y Procesar los datos en un sistema de información general basado en dashboard para la visualización de la información de la producción científica de las universidades

- Implementar los datos extraídos: Una vez finalizada la extracción, agrupamos todos los archivos en formato Json en una carpeta específica, luego importamos los metadatos a la herramienta visual Power Bi.
- Procesar los datos importados: para realizar este objetivo se requiere previamente haber exportado todos los metadatos de las instituciones licenciadas y además los datos adicionales obtenidos de la lista de instituciones licenciadas por SUNEDU al Power Bi, luego realizamos una relación entre ambas tablas mediante el atributo nombre de universidad, para obtener una data con información detalla de cada institución en la que se encuentra presente: su ubicación por departamento y provincia, fecha de licenciamiento y tipo de gestión.
- Identificar los indicadores clave de rendimiento relevantes: una vez que los datos estén disponibles en el Power Bi, usando la herramienta vista de datos determinamos los indicadores relevantes que podemos visualizar en el dashboard.

3.3.3 Objetivo específico 3: Evaluar el desempeño científico de los RI y plantear mejoras a partir de la información generada

- Segmentar los metadatos por macrorregiones y regiones naturales, para la evaluación del desempeño.
- Recopilar la información relevante y analizar los resultados.

CAPÍTULO IV RESULTADOS Y DISCUSIÓN

4.1 Resultado específico 1

Tabla 2

Cantidad de repositorios habilitados al momento de la validación

Tipo de Gestión	Repositorio Habilitado	Total de Repositorios
Privada	48	48
Públicas	44	47

Se determino que 3 de las 95 instituciones licenciadas por SUNEDU no contaban con el repositorio institucional habilitado al momento de realizar la validación.

Tabla 3

Cantidad de Repositorios que cuentan con SSL activado

Tipo de Gestión	SSL activado	SSL desactivado
Privada	40	8
Públicas	30	14
Total	70	22

Así mismo los RI habilitados contaban con la siguiente estructura de metadatos:

```
<record>
  <header>
    <identifier>oai:repositorio.unsm.edu.pe:11458/489</identifier>
    <timestamp>2022-10-11T17:57:03Z</timestamp>
    <setSpec>com_11458_2071</setSpec>
    <setSpec>com_11458_2003</setSpec>
    <setSpec>com_11458_1982</setSpec>
    <setSpec>com_11458_1979</setSpec>
    <setSpec>col_11458_2075</setSpec>
  </header>
  <metadata>
    <dc:dc xmlns:oai_dc="http://www.openarchives.org/OAI/2.0/oai_dc/" xmlns:dc="http://purl.org/dc/elements/1.1/" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xmlns:
    <dc:title>Relación entre el conocimiento de la problemática ambiental y la formación del pensamiento ecológico positivo de los alumnos del 5to grado de secundaria de las I
    <dc:creator>García Jimenez, Enrique</dc:creator>
    <dc:contributor>Vargas Saldaña, German</dc:contributor>
    <dc:subject>Problemática ambiental</dc:subject>
    <dc:subject>Pensamiento ecológico</dc:subject>
    <dc:subject>https://purl.org/pe-repo/ocde/ford#6.05.01</dc:subject>
    <dc:description>Already more than a decade ago has been inserted into the scope of the educational process the concept of learning achievement, the same meaning that proje
    <dc:description>Hace más de una década se ha insertado en el ámbito del proceso educativo el concepto de logro de aprendizaje, el mismo que proyecta su significado hacia l
    <dc:date>2016-11-02T19:10:19Z</dc:date>
    <dc:date>2016-11-02T19:10:19Z</dc:date>
    <dc:date>2012</dc:date>
    <dc:type>info:eu-repo/semantics/bachelorThesis</dc:type>
    <dc:identifier>García-Jimenez, E. (2012). Relación entre el conocimiento de la problemática ambiental y la formación del pensamiento ecológico positivo de los alumnos del
    <dc:identifier>http://hdl.handle.net/11458/489</dc:identifier>
    <dc:language>es</dc:language>
    <dc:rights>open</dc:rights>
    <dc:format>application/pdf</dc:format>
    <dc:publisher>Universidad Nacional de San Martín. Fondo Editorial</dc:publisher>
    <dc:publisher>PE</dc:publisher>
    <dc:source>Repositorio - UNSM</dc:source>
  </oai_dc:dc>
</metadata>
</record>
```

Figura 2.

Estructura de los metadatos

Una vez que tenemos los datos validados. Codificamos la extracción de los metadatos, utilizando métodos de web scraping en el lenguaje de programación Python.

Importamos las librerías necesarias:

- Requests: enviar HTTP solicitado usando Python y devuelve un objeto respuesta.
- Json: trabajar con datos en Json.

- BeautifulSoup: biblioteca en Python la cual permite analizar documentos HTML.
- Io: provee facilidades para manejar diferentes tipos de entradas y salidas.

```
import requests
import json
from bs4 import BeautifulSoup
import io
```

Luego indicamos las variables:

- url_set: la dirección Url donde se encuentran las etiquetas que categorizan al RI.
- url_content: la dirección Url donde se encuentran los metadatos del material científico de las universidades.
- var_nameuni: indicamos el nombre de la institución.
- var_namefile: indicamos el nombre del archivo Json.

```
url_set = 'https://repositorio.unsm.edu.pe/oai/request?verb=ListSets'
url_content=
'https://repositorio.unsm.edu.pe/oai/request?verb=ListRecords&metadataPrefix=oai_dc'
var_nameuni = 'Universidad Nacional de San Martín'
var_namefile = 'data_oai_unsm'
```

Extraemos los metadatos de la institución mostrada anteriormente en un archivo con formato Json el cual encaja perfectamente a la estructura de los metadatos variables que tiene cada institución, la extracción se realiza institución por institución luego de indicarle las variables necesarias.

Tabla 4

Cantidad de Repositorios que cuentan con SSL activado

Tipo de Gestión	Completo	Incompleto
Privada	41	7
Publicas	31	16
Total	72	23

Al finalizar el proceso de extracción identificamos que 23 de las 95 instituciones licenciadas no contaban con el oai habilitado o no presentaban el certificado ssl activado, la lista se puede ver en:

https://docs.google.com/spreadsheets/d/1Jo6WyZIWpMEV_aOJ3w4atYxhXpei-Rc6/edit?usp=sharing&oid=106909389732919702792&rtpof=true&sd=true.

Tabla 5

Cantidad de elementos extraídos

Nombre	Total
Universidad César Vallejo	108598
Universidad Peruana de Ciencias Aplicadas	50386

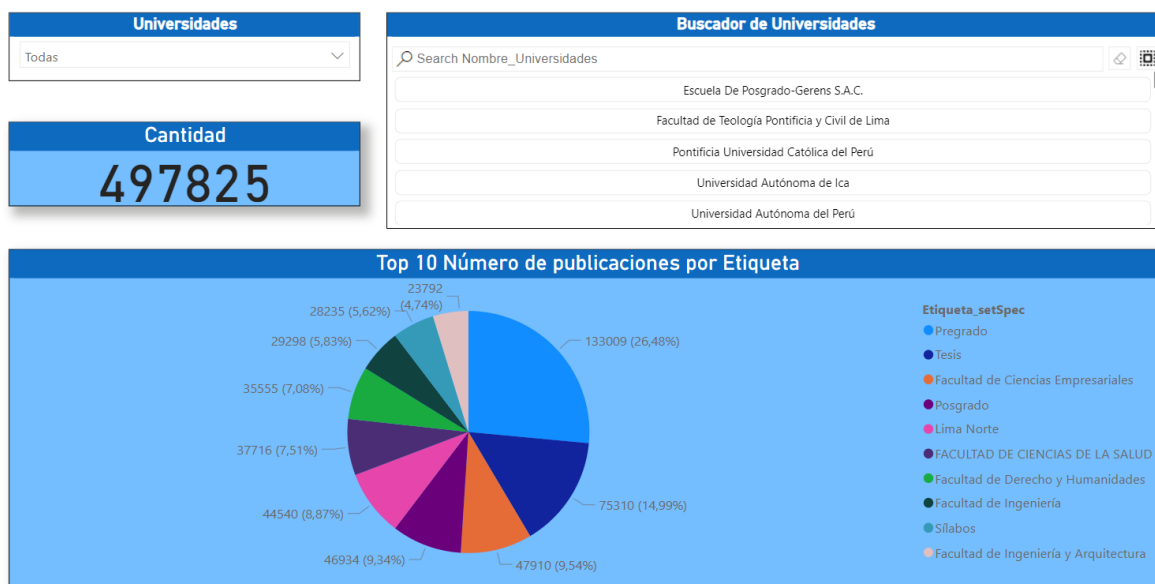
Nombre	Total
Universidad Nacional de Ingeniería	23953
Universidad San Ignacio de Loyola	22965
Pontificia Universidad Católica del Perú	21104
Universidad Nacional Mayor de San Marcos	18866
Universidad Privada del Norte	16016
Universidad Nacional de San Agustín	14100
Universidad Peruana Cayetano Heredia	12491
Universidad Continental	10537
Universidad Señor de Sipán	10529
Universidad de Lima	10488
Universidad de San Martín de Porres	10423
Universidad Nacional del Altiplano	9840
Universidad Católica de Santa María	9314
Universidad Privada Antenor Orrego	8576
Universidad Nacional del Centro del Perú	8023
Universidad Nacional José Faustino Sánchez Carrión	6874
Universidad Nacional de Educación Enrique Guzmán y Valle	6708
Universidad Nacional de la Amazonía Peruana	6562
Universidad Tecnológica del Perú	6498
Universidad Nacional de San Antonio Abad del Cusco	6039
Universidad Ricardo Palma	5519
Universidad Católica Santo Toribio de Mogrovejo	5515
Universidad Nacional de Cajamarca	5274
Universidad Peruana Los Andes	5122
Universidad de Piura	4438
Universidad Privada San Juan Bautista	4057
Universidad de Huánuco	3921
Universidad Nacional de San Martín	3788
Universidad Nacional de Piura	3742
Universidad Nacional Jorge Basadre Grohmann	3599
Universidad Nacional Santiago Antúnez de Mayolo	3430
Universidad del Pacífico	3357
Universidad Marcelino Champagnat	3311
Universidad Nacional de Ucayali	3267
Universidad Nacional de Tumbes	3057
Universidad Nacional de San Cristóbal de Huamanga	2956
Universidad Privada de Tacna	2729
Universidad Nacional Toribio Rodríguez de Mendoza de Amazonas	2705
Universidad Nacional Daniel Alcides Carrión	2699
Universidad Científica del Sur	2654
Universidad ESAN	2566
Universidad Católica San Pablo	2357
Universidad Nacional Agraria de la Selva	2191

Nombre	Total
Universidad Autónoma del Perú	2091
Universidad Autónoma de Ica	1850
Universidad Católica Sedes Sapientiae	1533
Universidad María Auxiliadora	1237
Universidad Privada de Huancayo Franklin Roosevelt	1155
Universidad Nacional San Luis Gonzaga	1154
Universidad Le Cordon Bleu S.A.C.	1062
Universidad Femenina del Sagrado Corazón	898
Universidad Nacional Tecnológica De Lima Sur	841
Universidad Nacional Micaela Bastidas de Apurímac	791
Universidad de Ciencias y Humanidades	750
Facultad de Teología Pontificia y Civil de Lima	671
Universidad Nacional José María Arguedas	547
Universidad Tecnológica de los Andes	434
Universidad de Ciencias y Artes de América Latina	237
Universidad Nacional de Juliaca	233
Universidad Nacional Autónoma de Chota	228
Universidad de Ingeniería y Tecnología	215
Universidad Jaime Bausate y Meza	199
Universidad Nacional de Frontera	181
Universidad Privada Norbert Wiener	100
Universidad Nacional de Moquegua	99
Universidad para el Desarrollo Andino	85
Universidad Nacional de Barranca	73
Escuela De Posgrado-Gerens S.A.C.	34
Universidad Nacional Intercultural de Quillabamba	24
Universidad Nacional Intercultural de la selva central Juan Santos Atahualpa	11

4.2 Resultado específico 2

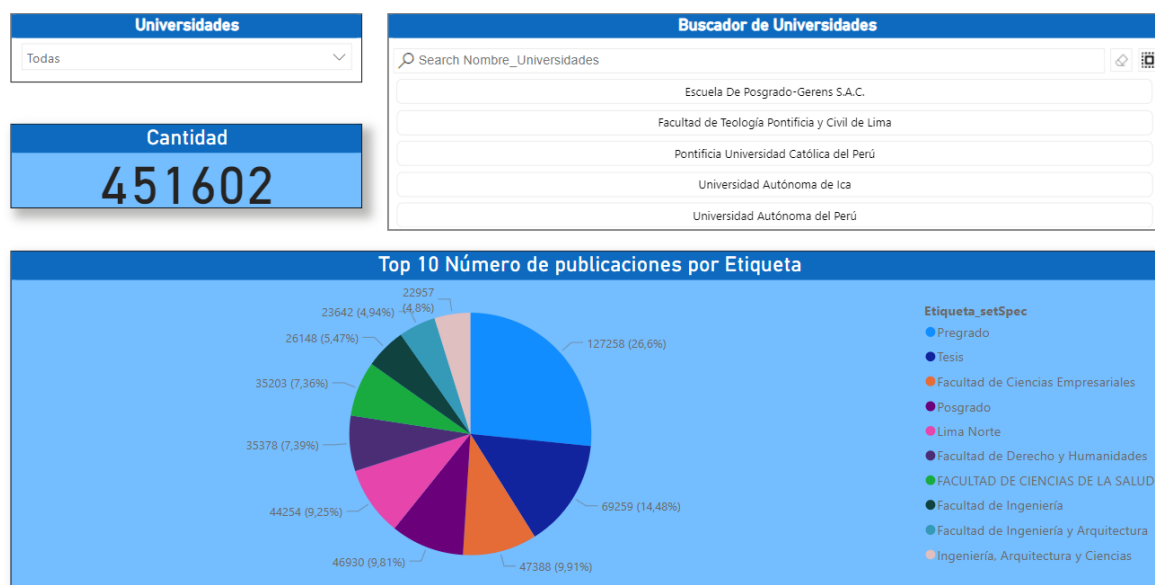
Para la implementación de los datos, es importante la importación de los archivos Json a la herramienta visual Power bi en conjunto, mediante el apartado obtener datos/ carpeta, pues al disponer de varios archivos Json, el apartado Json solo permite importar un archivo del mismo tipo. Adicional se importó los datos de la lista de instituciones licenciadas por SUNEDU la cual permite que se tenga los datos geográficos de las universidades.

Los filtros que agregamos al dashboard siguen los parámetros estipulados en la Guía de Alicia 2.0.1 en la que tomamos en cuenta: el tipo de acceso y el tipo de publicación válidos, al respetar este parámetro podemos determinar que muchas universidades no cumplen con el vocabulario o el tipo de campo necesario.

**Figura 3.**

Dashboard parte 1 sin filtros: tipo de dato y tipo de acceso

La cantidad total de elementos guardados en el conjunto de repositorios es de 497825 siendo estos solo de las instituciones licenciadas que tuvieron el oai habilitado al momento de la extracción.

**Figura 4.**

Dashboard parte 1 con filtros: tipo de dato y tipo de acceso

La cantidad total de elementos guardados en el conjunto de repositorios usando los filtros es de 451602, siendo 46223 la diferencia entre ambos.

Tabla 6

Total de elementos por tipo de etiqueta

Dashboard 1		Dashboard 2	
Nombre Universidades	Total	Nombre Universidades	Total
Pregrado	133009	Pregrado	127258
Tesis	75310	Tesis	69259

Facultad de Ciencias Empresariales	47910	Facultad de Ciencias Empresariales	47388
Posgrado	46934	Posgrado	46930
Lima Norte	44540	Lima Norte	44254
FACULTAD DE CIENCIAS DE LA SALUD	37716	Facultad de Derecho y Humanidades	35378
Facultad de Derecho y Humanidades	35555	FACULTAD DE CIENCIAS DE LA SALUD	35203
Facultad de Ingeniería	29298	Facultad de Ingeniería	26148
Sílabos	28235	Facultad de Ingeniería y Arquitectura	23642
Facultad de Ingeniería y Arquitectura	23792	Ingeniería, Arquitectura y Ciencia	22957

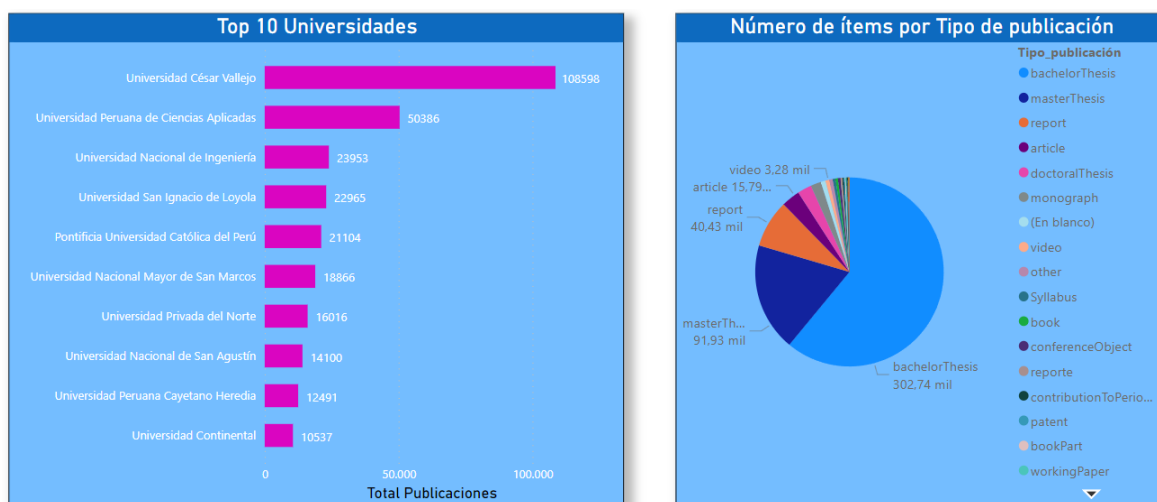


Figura 5.

Dashboard parte 2 sin filtro: Top 10 universidades y Tipo Publicación

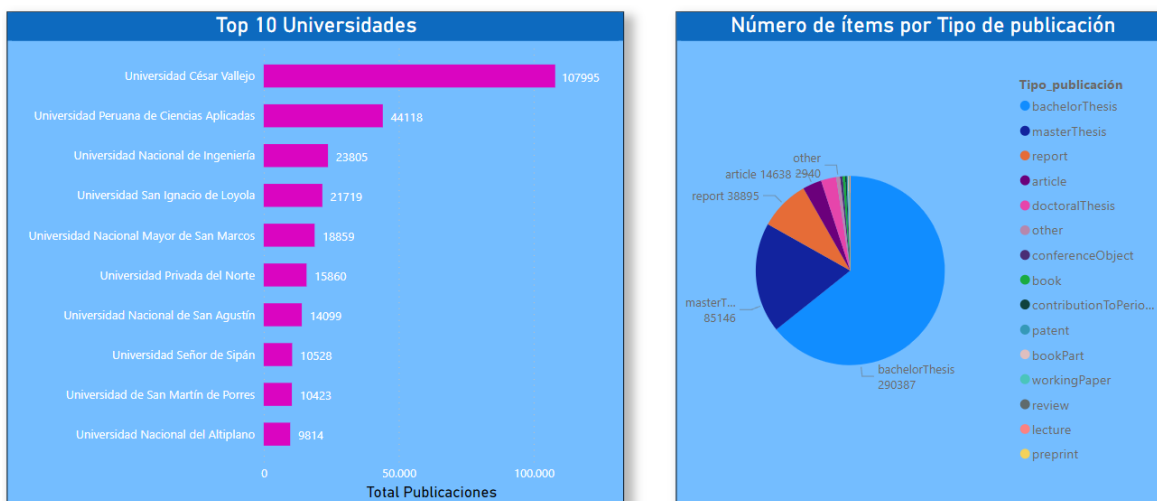


Figura 6.

Dashboard parte 2 con filtro: Top 10 universidades y Tipo Publicación

Tabla 7

Top 10 Universidades

Dashboard 1		Dashboard 2	
Nombre Universidades	Total	Nombre Universidades	Total
Universidad Cesar Vallejo	108598	Universidad Cesar Vallejo	107995
Universidad Peruana de Ciencias Aplicadas	50386	Universidad Peruana de Ciencias Aplicadas	44118

Universidad Nacional de Ingeniería	23953	Universidad Nacional de Ingeniería	23805
Universidad San Ignacio de Loyola	22965	Universidad San Ignacio de Loyola	21719
Pontificia Universidad Católica del Perú	21104	Universidad Nacional Mayor de San Marcos	18859
Universidad Nacional Mayor de San Marcos	18866	Universidad Privada del Norte	15860
Universidad Privada del Norte	16016	Universidad Nacional de san Agustín	14099
Universidad Nacional de san Agustín	14100	Universidad Señor de Sipán	10528
Universidad Peruana Cayetano Heredia	12791	Universidad San Martín de Porres	10423
Universidad Continental	10537	Universidad Nacional del Antiplano	9814

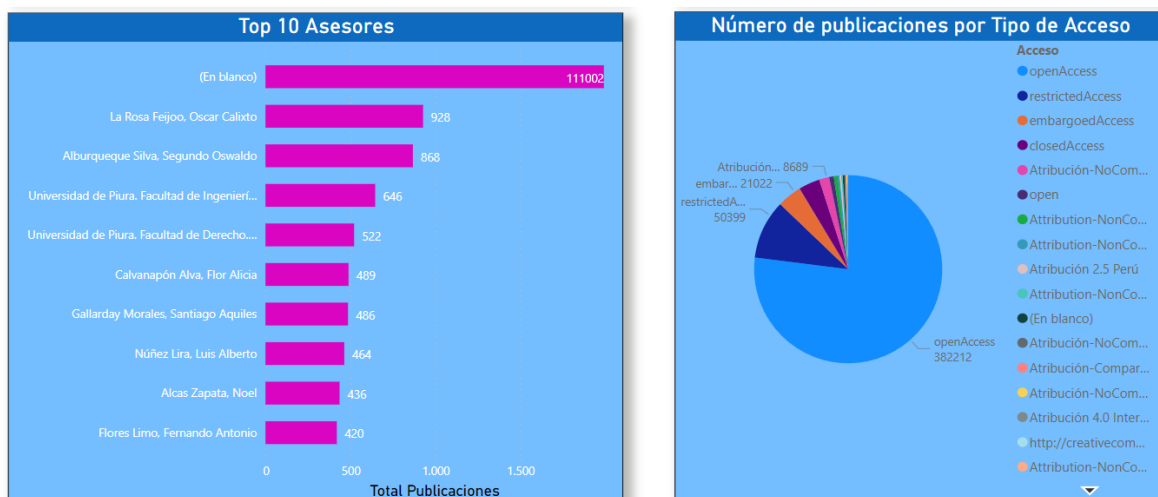
En la tabla 7 podemos ver que en ambos casos la universidad Cesar Vallejo mantiene el primer lugar por la gran cantidad de elementos registrados en su repositorio institucional.

Tabla 8

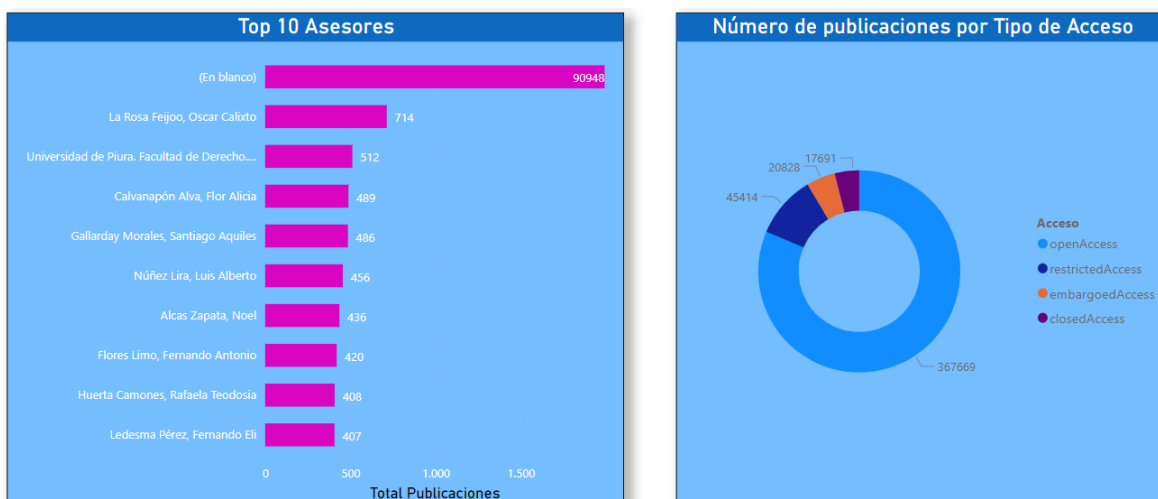
Número de elementos por tipo de publicación

Dashboard 1		Dashboard 2	
Nombre Universidades	Total	Nombre Universidades	Total
bachelorThesis	302743	bachelorThesis	290387
masterThesis	97956	masterThesis	85146
report	40430	report	38895
artcile	15785	artcile	14638
doctoralThesis	12069	doctoralThesis	11771
monograph	8855	other	2940
(En blanco)	3986	confereneObjet	2044
Video	3279	book	1990
other	2977	contributionToPeriodical	1056
Syllabus	2338	patent	919
book	2125	bookPart	655
conferenceObject	2078	workingPaper	604
reporte	1695	review	469
contributionToPeriodical	1099	lecture	85
patent	919	preprint	54

En la tabla 8 ambos dashboard se puede identificar que bachelorThesis y masterThesis poseen la mayor cantidad de elementos, siendo estos determinados como tesis de bachiller y tesis de maestría.

**Figura 7.**

Dashboard parte 3 sin filtros: tipo de dato y tipo de acceso

**Figura 8.**

Dashboard parte 3 con filtros: tipo de dato y tipo de acceso

Tabla 9*Top 10 Asesores agrupación sin filtro*

Asesor	Universidad	Total
(En Blanco)	Escuela de Posgrado-Gerens S.A.C	111002
La Rosa Feijoo, Oscar Calixto	Universidad Nacional de Tumbes	928
Albuquerque silva, Segundo Oswaldo	Universidad Nacional de Tumbes	868
Universidad de Piura Facultad de Ingeniería Departamento de ingeniería Industrial y Sistemas	Universidad de Piura	646
Universidad de Piura. Facultad de Derecho Área Departamental de Derecho	Universidad de Piura	522
Calvanapón Alva, Flor Alicia	Universidad César Vallejo	489
Gallarday Morales, Santiago Aquiles	Universidad César Vallejo	486
Núñez Lira, Luis Alberto	Universidad César Vallejo	464
Alcas Zapata, Noel	Universidad César Vallejo	436
Flores Limo, Fernando Antonio	Universidad César Vallejo	420

Tabla 10*Top 10 Asesores agrupación con filtro*

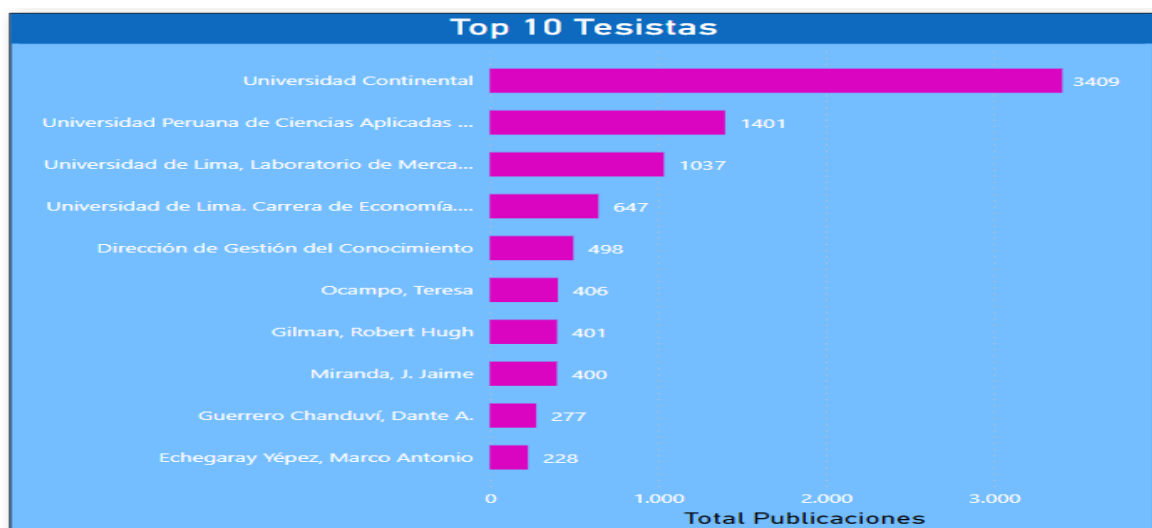
Asesor	Universidad	Total
(En Blanco)	Escuela de Posgrado-Gerens S.A.C	90948
La Rosa Feijoo, Oscar Calixto	Universidad Nacional de Tumbes	714
Universidad de Piura. Facultad de Derecho. Área Departamental de Derecho	Universidad de Piura	512
Calvanapón Alva, Flor Alicia	Universidad César Vallejo	489
Gallarday Morales, Santiago Aquiles	Universidad César Vallejo	486
Núñez Lira, Luis Alberto	Universidad César Vallejo	456
Alcas Zapata, Noel	Universidad César Vallejo	436
Flores Limo, Fernando Antonio	Universidad César Vallejo	420
Huerta Camones, Rafaela Teodosia	Universidad Nacional de Educación Enrique Guzmán y Valle	408
Ledesma Pérez, Fernando Eli	Universidad César Vallejo	407

En la tabla 9 y tabla 10, se determina que una gran cantidad de materia científico no cuenta con un asesor, siendo registrada en la Escuela de Posgrado-Gerens S.A.C.

Tabla 11*Comparación de Número de publicaciones por Tipo de Acceso*

Dashboard 1		Dashboard 2	
Tipo de Acceso	Total	Tipo de Acceso	Total
openAccess	382212	openAccess	367669
restrictedAccess	50399	restrictedAccess	45414
embargoedAccess	21022	embargoedAccess	20828
closedAcces	18008	closedAcces	17691

En la Tabla 11 la cual está relacionada al dashboard 1 de la Figura 7 identificamos que el conjunto de universidades mantiene una variedad adicional de tipos de accesos las cuales no se encuentran estipuladas en la Guía de Alicia.

**Figura 9.**

Dashboard parte 4 sin filtros: Top 10 Tesistas

Tabla 12*Top 10 Tesistas sin filtros*

Autor	Universidad	Total
Universidad Continental	Universidad Continental	3409
Universidad Peruana de Ciencias Aplicadas UPC	Universidad Peruana de Ciencias Aplicadas	1401
Universidad de Lima, Laboratorio de Mercado de Capitales	Universidad de Lima	1037
Universidad de Lima. Carrera de Economía. Laboratorio de Mercado de Capitales	Universidad de Lima	647
Dirección de Gestión del Conocimiento	Universidad Peruana de Ciencias Aplicadas	498
Ocampo, Teresa	Universidad Le Cordon Bleu S.A.C	406
Gilman, Robert Hugh	Universidad Peruana Cayetano Heredia	401
Miranda, J. Jaime	Universidad de Lima	400
Guerrero Chanduvi, Dante A.	Universidad de Piura	277
Echegaray Yépez, Marco Antonio	Universidad Continental	228

**Figura 10.**

Dashboard parte 4 con filtros: Top 10 Tesistas

Tabla 13*Top 10 Tesistas con filtros*

Autor	Universidad	Total
Universidad de Lima, Laboratorio de Mercado de Capitales	Universidad de Lima	1011
Universidad Peruana de Ciencias Aplicadas UPC	Universidad Peruana de Ciencias Aplicadas	735
Universidad de Lima. Carrera de Economía. Laboratorio de Mercado de Capitales	Universidad de Lima	557
Universidad Continental	Universidad Continental	413
Ocampo, Teresa	Universidad Le Cordon Bleu S.A.C	406

Gilman, Robert Hugh	Universidad Peruana Cayetano Heredia	400
Miranda, J. Jaime	Universidad de Lima	398
Dirección de Gestión del Conocimiento	Universidad Peruana de Ciencias Aplicadas	351
Bernabé Ortiz, Antonio	Universidad de Lima	226
Castillo Córdova, Luis	Universidad de Piura	136

En la tabla 12 y 13 podemos identificar que el autor en varios elementos lleva el nombre de la universidad a la que está vinculada, siendo esto posible a la mala gestión de metadatos o también a los distintos tipos de publicación a los cuales están relacionados.

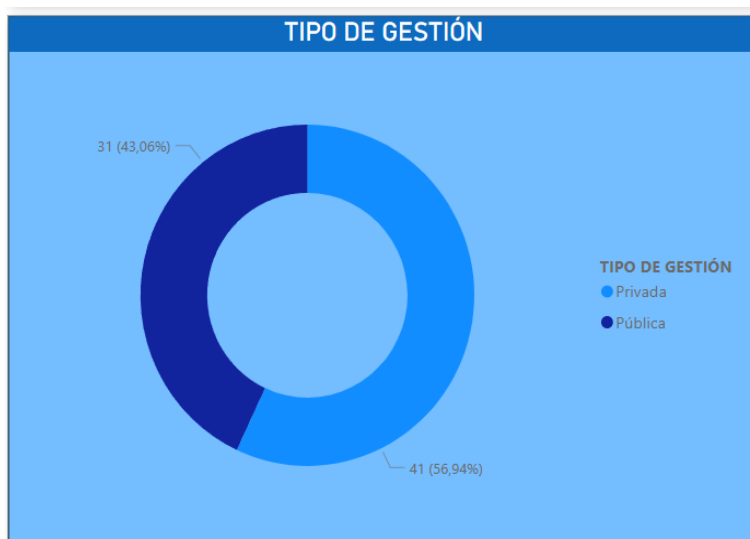


Figura 11.
Dashboard parte 5 sin filtros: tipo de gestión

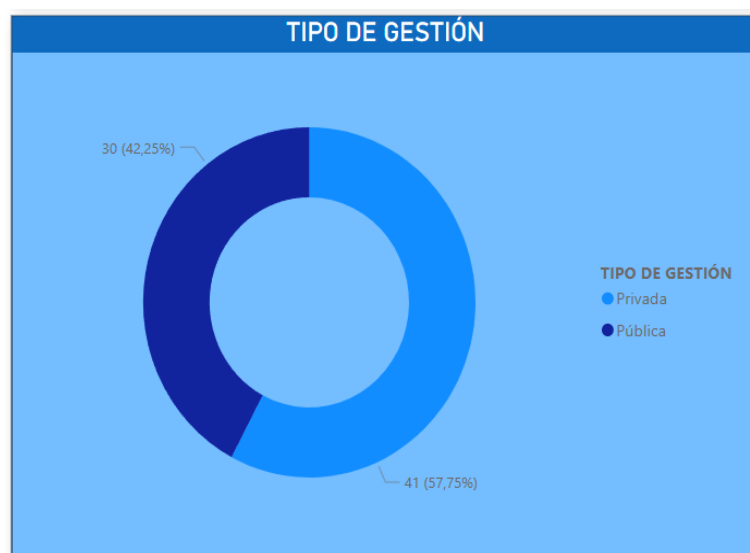
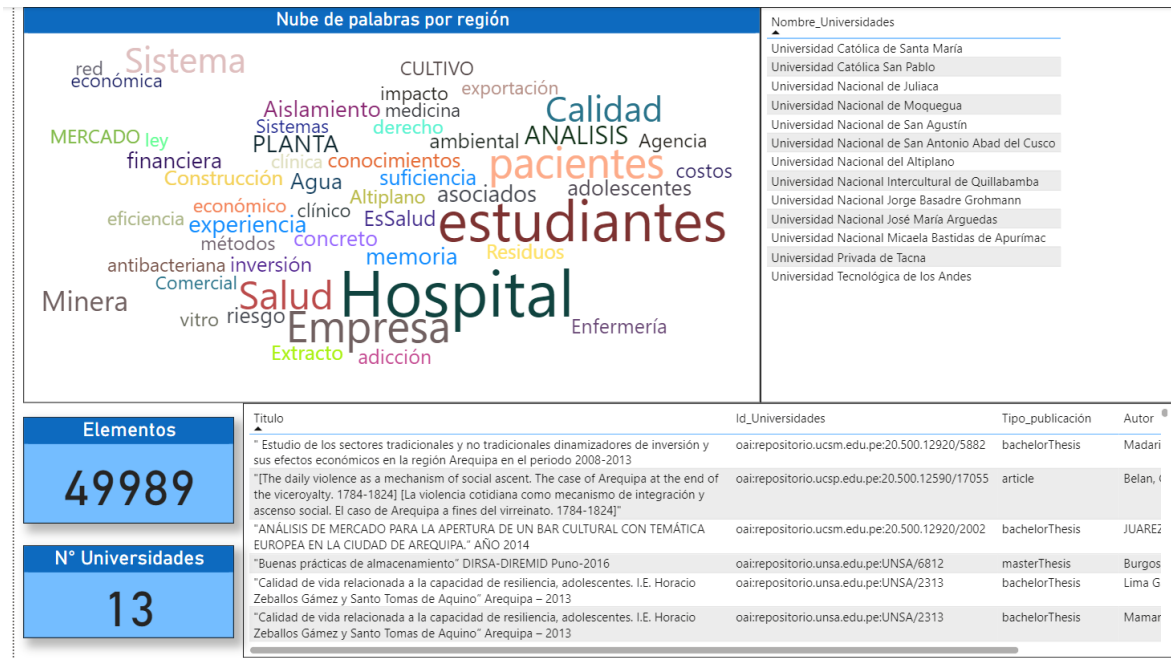


Figura 12.
Dashboard parte 5 con filtros: tipo de gestión

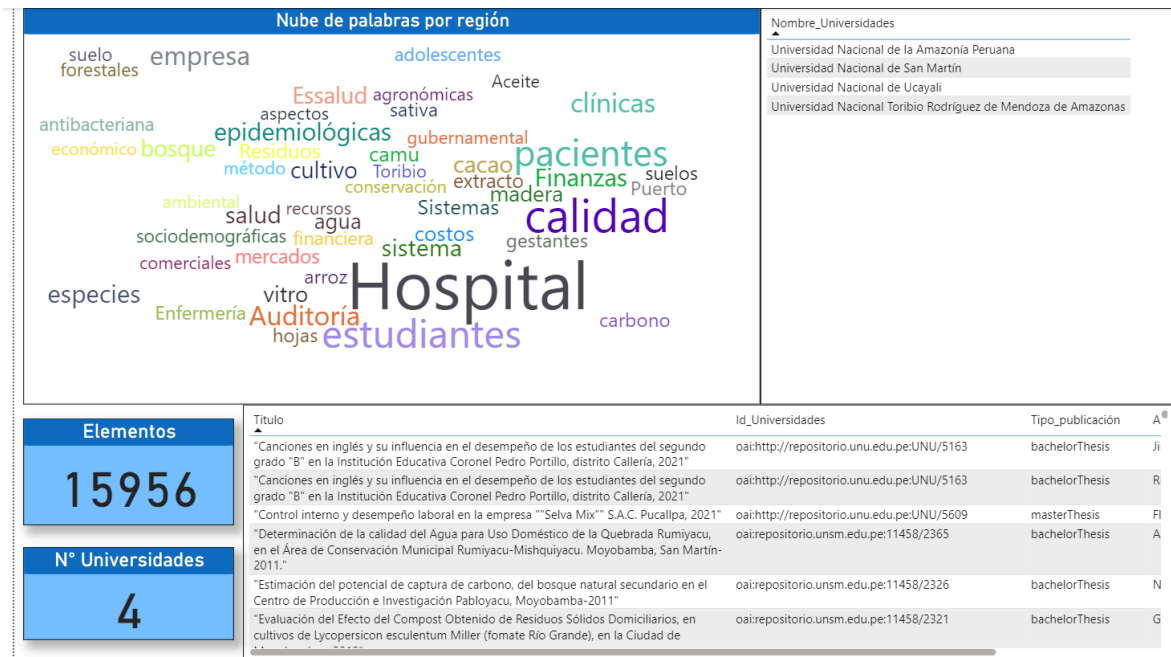
En la figura 11 y 12 podemos determinar que el mayor porcentaje de universidades con el oai habilitado es de tipo gestión privada.



Marcador 5 de 7

Sur

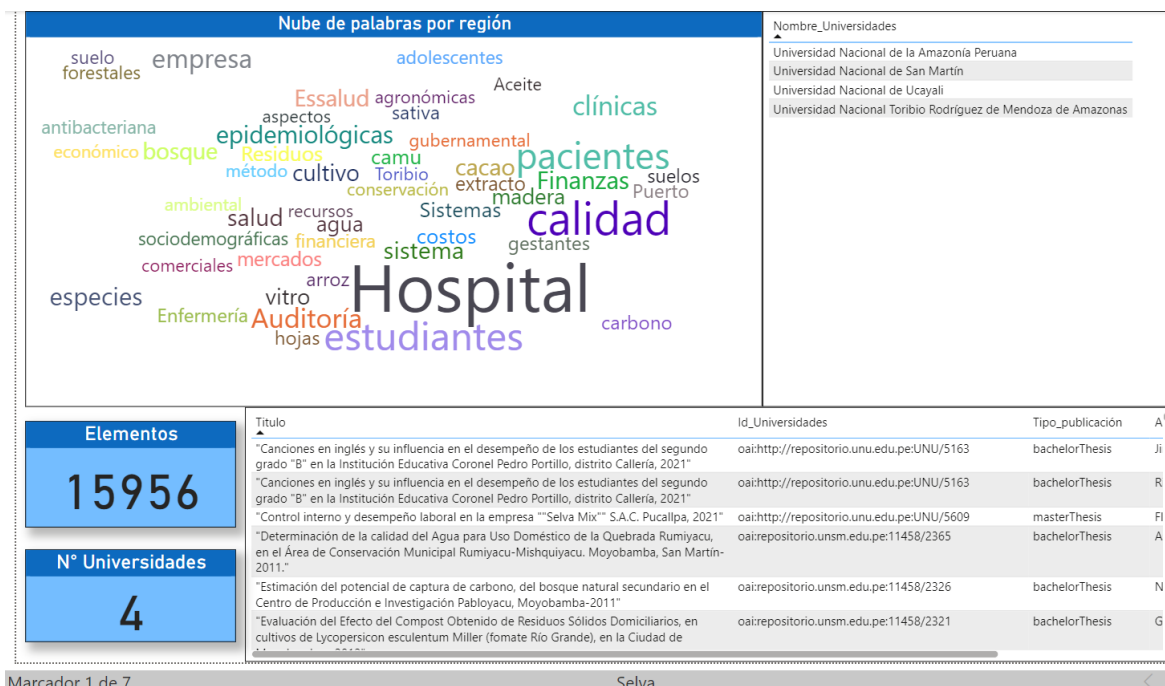
Figura 15.
Macrorregión Sur



Marcador 4 de 7

Oriente

Figura 16.
Macrorregión Oriente



Marcador 1 de 7
Figura 19.
 Región Selva

Tabla 14
Cantidad de elementos por Región

Macrorregiones	Cantidad de elementos	N° de Universidades	Razón de elementos por universidad
Norte	160,906	11	14,627
Centro	226,417	43	5,262
Sur	49,989	13	3,845
Oriente	15,956	4	3,989
Regiones Geográficas	Cantidad de elementos	N° de Universidades	
Costa	388,461	49	7,929
Sierra	51,048	19	2,687
Selva	15,956	4	3,989

La tabla 14 muestra que, en la región norte y centro, así como la costa del país se encuentra la mayor cantidad de material académico. Además, al calcular la razón de elementos (material científico) la cual se calcula con: **Razón de material científico por universidad = Cantidad de material científico / Número de universidades**. Podemos determinar que un valor más alto, indica que hay un mejor acceso a recursos científicos y académicos.

Adicionalmente determinamos que el top 5 de palabras más relevantes:

Por macrorregiones del Perú

1. Norte: Empresa, estudiantes, productividad, calidad, sistema.
2. Centro: Estudiantes, comprensión, matemática, interaprendizaje, aprendizaje.

3. Sur: Hospital, estudiantes, empresa, pacientes, salud.
4. Oriente: Hospital, calidad, estudiantes, pacientes, auditoría.

Por regiones geográficas del Perú

1. Costa: Estudiantes, comprensión, matemática, interaprendizaje, aprendizaje.
2. Sierra: Estudiantes, hospital, salud, sistema, adolescentes.
3. Selva: Hospital, calidad, estudiantes, pacientes, auditoría.

Comparando el top 5 de palabras más relevantes podemos determinar que el apartado médico, empresarial y de educación son las áreas más concurrentes para la investigación en la actualidad.

CONCLUSIONES

1. Logramos realizar una extracción correcta de los metadatos mediante un algoritmo de web scraping usando el protocolo OAI habilitado en los repositorios institucionales. Luego de evaluar la extracción de datos identificamos que un grupo de universidades no contaban con oai habilitado al momento de la extracción realizada el 20 de mayo del 2023.
2. Agregamos exitosamente los metadatos de las universidades licenciadas a un sistema de información basado en dashboard, tras el análisis de la estructura que presentaban los metadatos, se identificó que muchas universidades no respetan el vocabulario asignado para los metadatos.
3. Mediante el procesamiento de datos y ayuda de la visualización del dashboard identificamos que un porcentaje mayor de universidades licenciadas que presentan un repositorio institucional habilitado, pertenecen al tipo de gestión privada.
4. Determinamos que la gran cantidad de datos limita el análisis del desempeño científico de las universidades peruanas, puesto que la visualización de nube de palabras solo admite una cantidad de 30000 elementos, haciendo que se considere viable una solución enfocada en Big Data o mediante aplicaciones especializadas para un volumen muy grande de datos.

RECOMENDACIONES

1. Para futuras investigaciones relacionadas al tema, recomendamos tener en cuenta el volumen de datos el cual llega a ser una limitante para las gráficas en la evaluación de los datos.
2. Recomendamos para futuras investigaciones, las cuales requieran trabajar con los metadatos, la utilización y evaluación de diferentes opciones de limpieza de datos, considerando un enfoque a Big Data.
3. Con el fin de automatizar el proceso de extracción y procesamiento de datos, se recomienda considerar soluciones en la nube, las cuales podrían permitir mayor capacidad a la investigación.
4. Recomendamos a los encargados de la administración de metadatos codificar el valor de la fecha de acuerdo a la norma ISO 8601 como establece la Guía de Alicia, con el formato de (año, mes, día) o (año y mes) para tener mayor alcance en la evaluación del material científico de las universidades.

REFERENCIAS BIBLIOGRÁFICAS

- ANDRÉS, G., AMADIO, C.C., WURSTEN, A. y LEGARIA, J.I., 2020. Open access institutional digital repositories: A socio-technical approach. A case study of the repository of the universidad nacional de Entre Rios-Argentina. *Informacion, Cultura y Sociedad*, no. 42, ISSN 18511740. DOI 10.34096/ICS.I42.7404.
- ÁVILA-CHAVES, G.A., 2021. *Diseño de un dashboard de control de la operación del equipo de respaldos Caso: GBM* [en línea]. S.l.: s.n. [consulta: 13 julio 2022]. Disponible en: <https://repositoriotec.tec.ac.cr/handle/2238/13232>.
- BAMIGBOLA, A.A. y ADETIMIRIN, A.E., 2020. Assessing Determinants of Perceived Ease of Use of Institutional Repositories by Lecturers in Nigerian Universities. *International Information & Library Review* [en línea], vol. 52, no. 2, [consulta: 9 junio 2022]. ISSN 1057-2317. DOI 10.1080/10572317.2019.1662261. Disponible en: <https://www.tandfonline.com/doi/full/10.1080/10572317.2019.1662261>.
- BAUDOIN, P. y BRANSCHOFKY, M., 2008. Implementing an Institutional Repository. https://doi.org/10.1300/J122v24n01_04 [en línea], [consulta: 25 junio 2022]. ISSN 0194262X. DOI 10.1300/J122V24N01_04. Disponible en: https://www.tandfonline.com/doi/abs/10.1300/J122v24n01_04.
- BLANCO-OLEA, F.S., 2021a. Comparative quality evaluation of universities' institutional repositories of peru. *JLIS.it*, vol. 12, no. 2, ISSN 20381026. DOI 10.4403/JLIS.IT-12688.
- BLANCO-OLEA, F.S., 2021b. Comparative quality evaluation of universities' institutional repositories of peru. *JLIS.it*, vol. 12, no. 2, ISSN 20381026. DOI 10.4403/jlis.it-12688.
- BRÉANT, C., SUCCI, L., COTTEN, M., GRIMAUD, S., IAVINDRASANA, J., KINDSTRAND, M., MAUVAIS, F. y RORIVE-FEYTMANS, B., 2020. Tools to measure, monitor, and analyse the performance of the Geneva university hospitals (HUG). <https://doi.org/10.1080/16258312.2020.1780634> [en línea], vol. 21, no. 2, [consulta: 24 junio 2022]. ISSN 16246039. DOI

- 10.1080/16258312.2020.1780634. Disponible en:
<https://www.tandfonline.com/doi/abs/10.1080/16258312.2020.1780634>.
- CHISITA, C.T. y CHIPARAUSHA, B., 2021. An Institutional Repository in a Developing Country: security and Ethical Encounters at the Bindura University of Science Education, Zimbabwe. *New Review of Academic Librarianship*, vol. 27, no. 1, ISSN 17407834. DOI 10.1080/13614533.2020.1824925.
- CONEJERO, J.M., PRECIADO, J.C., FERNÁNDEZ-GARCÍA, A.J., PRIETO, A.E. y RODRÍGUEZ-ECHEVERRÍA, R., 2021. Towards the use of Data Engineering, Advanced Visualization techniques and Association Rules to support knowledge discovery for public policies. *Expert Systems with Applications*, vol. 170, ISSN 09574174. DOI 10.1016/J.ESWA.2020.114509.
- CÓRDOVA VIERA, Y., MARTÍNEZ BORREGO, J. y CÓRDOVA VIERA, E., 2021. Propuesta de metodología para el diseño de dashboard. [en línea], [consulta: 27 junio 2022]. DOI 10.5281/ZENODO.5545998. Disponible en: <https://zenodo.org/record/5545998>.
- DOBRAJA, I., MENNO, &, KRAAK, J. y KRAAK, M.-J., 2020. Principles of dashboard adaptability to get insights into origin-destination data. <https://doi.org/10.1080/17489725.2020.1738577> [en línea], vol. 14, no. 1, [consulta: 2 junio 2022]. ISSN 17489733. DOI 10.1080/17489725.2020.1738577. Disponible en: <https://www.tandfonline.com/doi/abs/10.1080/17489725.2020.1738577>.
- GHOSH, S. y ROY, B.K., 2021. Webometric analysis of open access digital repository of agricultural sciences in Asia. <https://doi.org/10.1080/09737766.2021.1938743> [en línea], vol. 15, no. 1, [consulta: 21 junio 2022]. ISSN 0973-7766. DOI 10.1080/09737766.2021.1938743. Disponible en: <https://www.tandfonline.com/doi/abs/10.1080/09737766.2021.1938743>.
- GREENBERG, J., 2009. Understanding Metadata and Metadata Schemes. https://doi.org/10.1300/J104v40n03_02 [en línea], vol. 40, no. 3-4, [consulta:

- 25 junio 2022]. ISSN 15444554. DOI 10.1300/J104V40N03_02. Disponible en: https://www.tandfonline.com/doi/abs/10.1300/J104v40n03_02.
- GUNDERMAN, H.C., SCHERER, D. y BEHRMAN, K., 2021. Leveraging library technology resources for internal projects, outreach, and engagement: A case study of Overleaf, LaTeX, and the KiltHub Institutional Repository Service at Carnegie Mellon University Libraries. *College and Undergraduate Libraries*, vol. 27, no. 2-4, ISSN 15452530. DOI 10.1080/10691316.2021.1885549.
- KENIGSBERG, T.A., HAUSE, A.M., MCNEIL, M.M., NELSON, J.C., ANN SHOUP, J., GODDARD, K., LOU, Y., HANSON, K.E., GLENN, S.C. y WEINTRAUB, E.S., 2022. Dashboard development for near real-time visualization of COVID-19 vaccine safety surveillance data in the vaccine safety datalink. *Vaccine*, vol. 40, no. 22, ISSN 18732518. DOI 10.1016/J.VACCINE.2022.04.010.
- LOCK, O., BEDNARZ, T. y PETTIT, C., 2020. The visual analytics of big, open public transport data – a framework and pipeline for monitoring system performance in Greater Sydney. <https://doi.org/10.1080/20964471.2020.1758537> [en línea], vol. 5, no. 1, [consulta: 23 junio 2022]. ISSN 25745417. DOI 10.1080/20964471.2020.1758537. Disponible en: <https://www.tandfonline.com/doi/abs/10.1080/20964471.2020.1758537>.
- MASINDE, J.M. y SANYA, O., 2022. Analysis of Interoperability, Security and Usability of Digital Repositories in Kenyan Institutions of Higher Learning. *Data and Information Management*, ISSN 2543-9251. DOI 10.1016/J.DIM.2022.100011.
- MERING, M., 2019. Transforming the quality of metadata in institutional repositories. *Serials Librarian*, vol. 76, no. 1-4, ISSN 15411095. DOI 10.1080/0361526X.2019.1540270.
- PRICE, R. y MURTAGH, J., 2020. An Institutional Repository Publishing Model for Imperial College London Grey Literature. *Serials Librarian*, vol. 79, no. 3-4, ISSN 15411095. DOI 10.1080/0361526X.2020.1847737.

- RANDELL, R., ALVARADO, N., ELSHEHALY, M., MCVEY, L., WEST, R.M., DOHERTY, P., DOWDING, D., FARRIN, A.J., FELTBOWER, R.G., GALE, C.P., GREENHALGH, J., LAKE, J., MAMAS, M., WALWYN, R. y RUDDLE, R.A., 2022. Design and evaluation of an interactive quality dashboard for national clinical audit data: a realist evaluation. *Health and Social Care Delivery Research*, vol. 10, no. 12, ISSN 2755-0060. DOI 10.3310/WBKW4927.
- SAINATH DANDAWATE, V. y DHANMJAYA, M., 2021. Open Access Initiatives in Western Asia. *Slavic and East European Information Resources*, vol. 22, no. 1, ISSN 15229041. DOI 10.1080/15228886.2021.1874349.
- SAN GIL, I., HUTCHISON, V., FRAME, M. y PALANISAMY, G., 2010. Metadata Activities in Biology. <https://doi.org/10.1080/19386389.2010.506389> [en línea], vol. 10, no. 2-3, [consulta: 25 junio 2022]. ISSN 19386389. DOI 10.1080/19386389.2010.506389. Disponible en: <https://www.tandfonline.com/doi/abs/10.1080/19386389.2010.506389>.
- SANCHEZ, M.E. y MARIÑO, S.I., 2021. Implementación de un repositorio para apoyo a la gestión administrativa. *Palabra clave (La Plata)* [en línea], vol. 10, no. 2, [consulta: 24 junio 2022]. ISSN 1853-9912. DOI 10.24215/18539912E130. Disponible en: <https://www.palabraclave.fahce.unlp.edu.ar/article/view/PCe130/13751>.
- SCHOEN, R.C., LAVENIA, M., CHICKEN, E., RAZZOUK, R., KISA, Z. y BOYLAN, M., 2019. Increasing secondary-level teachers' knowledge in statistics and probability: Results from a randomized controlled trial of a professional development program. *Cogent Education*, vol. 6, no. 1, ISSN 2331186X. DOI 10.1080/2331186X.2019.1613799.
- SCOTT, R.E., 2022. Data Scraping YouTube for the Study of Lieder Reception. *Nineteenth-Century Music Review*, vol. 19, no. 3, ISSN 20448414. DOI 10.1017/S1479409822000143.
- SHELLEY, A., 2020. It Takes a Village: Populating the Institutional Repository with Performing Arts Content. *Music Reference Services Quarterly*, vol. 23, no. 3-4, ISSN 15409503. DOI 10.1080/10588167.2020.1786308.

- SUSNJAK, T., RAMASWAMI, G.S. y MATHRANI, A., 2022. Learning analytics dashboard: a tool for providing actionable insights to learnerSusnjak, T., Ramaswami, G. S., & Mathrani, A. (2022). Learning analytics dashboard: a tool for providing actionable insights to learners. *International Journal of Educational Technology in Higher Education*, vol. 19, no. 1, ISSN 23659440. DOI 10.1186/S41239-021-00313-7.
- TURENNE, N., 2023. Net activism and whistleblowing on YouTube: a text mining analysis. *Multimedia Tools and Applications*, vol. 82, no. 6, ISSN 15737721. DOI 10.1007/S11042-022-13777-0.
- VARDELL, E., 2020. Global Health Observatory Data Repository. *Medical Reference Services Quarterly*, vol. 39, no. 1, ISSN 15409597. DOI 10.1080/02763869.2019.1693231.
- VÁZQUEZ-INGELMO, A., GARCÍA-PEÑALVO, F.J. y THERÓN, R., 2021. Towards a technological ecosystem to provide information dashboards as a service: A dynamic proposal for supplying dashboards adapted to specific scenarios. *Applied Sciences (Switzerland)*, vol. 11, no. 7, ISSN 20763417. DOI 10.3390/APP11073249.
- WANG, S., LI, G., YU, W. y MA, Y., 2021. Requirement-driven remote sensing metadata planning and online acquisition method for large-scale heterogeneous data. <http://www.tandfonline.com/action/journalInformation?show=aimsScope&journalCode=tgsi20#.VsXpLiCLRhE> [en línea], [consulta: 25 junio 2022]. ISSN 10095020. DOI 10.1080/10095020.2021.1994358. Disponible en: <https://www.tandfonline.com/doi/abs/10.1080/10095020.2021.1994358>.
- WHITE, P.B. y RADIO, E., 2020. Mapping Search Queries to Metadata Fields in a GeoBlacklight Repository. *Journal of Library Metadata*, vol. 20, no. 4, ISSN 19375034. DOI 10.1080/19386389.2020.1915459.
- WONG, E.Y., 2019. From Digital Commons to Scholar Profiles: Implementing a New System to Raise College Academic Distinction. <https://doi.org/10.1080/07317131.2019.1691759> [en línea], vol. 37, no. 1,

[consulta: 27 junio 2022]. ISSN 07317131. DOI 10.1080/07317131.2019.1691759. Disponible en: <https://www.tandfonline.com/doi/abs/10.1080/07317131.2019.1691759>.

WU, X., DINÇER, H. y YÜKSEL, S., 2022. Analysis of crowdfunding platforms for microgrid project investors via a q-rung orthopair fuzzy hybrid decision-making approach. *Financial Innovation*, vol. 8, no. 1, ISSN 2199-4730. DOI 10.1186/S40854-022-00353-6.

Agregación de repositorios institucionales para la generación de información del desempeño científico de universidades peruanas

por Edy Luis Huancaruna Golac

Fecha de entrega: 24-nov-2023 09:37a.m. (UTC-0500)

Identificador de la entrega: 2237442628

Nombre del archivo: Edy_Luis_Huancaruna_Golac_2.docx (2.46M)

Total de palabras: 9518

Total de caracteres: 56381

Agregación de repositorios institucionales para la generación de información del desempeño científico de universidades peruanas

INFORME DE ORIGINALIDAD



FUENTES PRIMARIAS

1	repositorio.unsm.edu.pe Fuente de Internet	5%
2	Miguel Valles-Coral, Richard Injante, Edwin Hernández-Torres, Lloy Pinedo et al. "Agregación de repositorios institucionales para la generación de información del desempeño científico de universidades peruanas", Iberoamerican Journal of Science Measurement and Communication, 2023 Publicación	4%
3	tesis.unsm.edu.pe Fuente de Internet	3%
4	www.adscientificindex.com Fuente de Internet	2%
5	alicia.concytec.gob.pe Fuente de Internet	1%
6	www.pcm.gob.pe Fuente de Internet	<1%