



e.ciencia

Encuentro Latinoamericano

2 0 1 9

Actas  
3er Encuentro  
Latinoamericano  
de e-Ciencia

Cancun International Convention Center  
(Cancun ICC), 2 - 4 septiembre, 2019  
Cancún, México





**Actas**  
**3<sup>er</sup> Encuentro Latinoamericano de e-Ciencia**

**“El genoma estudiantil y la  
metamorfosis digital  
universitaria”**

Cancun International Convention Center (Cancun ICC), 2 – 4 septiembre, 2019  
Cancún, México

**Comité de Programa:**

- ERNESTO CHINKES  
Presidente Honorario, Argentina
- LEANDRO GUIMARÃES  
Director Adjunto Escuela Superior de Redes, RNP,  
Brasil
- ALONSO CASTRO  
Director del Centro de Informática, Universidad de  
Costa Rica, Costa Rica
- LUIS A. NÚÑEZ  
Físico, Universidad Industrial de Santander,  
Bucaramanga - Colombia
- LUIS ALBERTO GUTIÉRREZ DÍAZ DE LEÓN  
Coordinador General de Tecnologías de Información  
Universidad de Guadalajara, México

Compiladora, coordinadora general de la publicación, edición y diseño:  
María José López Pourailly, RedCLARA.

RedCLARA (<http://www.redclara.net>)

Fecha en que se terminó la presente edición: 08-10-2019

ISBN: En trámite

Copyright de la presente edición:

Actas 3<sup>er</sup> Encuentro Latinoamericano de e-Ciencia  
“El genoma estudiantil y la metamorfosis digital universitaria”  
por RedCLARA,  
se encuentra bajo una Licencia Creative Commons  
Atribución-NoComercial-SinDerivadas 3.0 Unported.

## **Índice**

<b>Capítulo 1, Ciencia Abierta</b>	<b>7</b>
Impacto del proyecto ELINF en el acceso abierto dentro de la Educación Superior Cubana	9
Repositorio Institucional del Centro de Ciencias de la Atmósfera, UNAM	23
Diseño de interfaces para entornos de interacción performáticos distribuidos de creación sonora y visual	38
<b>Capítulo 2, Analítica de datos e inteligencia artificial en la Ciencia</b>	<b>55</b>
Implementación de un sistema experto basado en inteligencia computacional híbrida para ayuda diagnóstica en la especialidad de psiquiatría en pacientes que sufren depresión	57
Uso de metaheurísticas en la segmentación de mamografías para el prediagnóstico de cáncer de mama	68



## **Capítulo 1**

### **Ciencia Abierta**



## Impacto del proyecto ELINF en el acceso abierto dentro de la Educación Superior Cubana

Dr.C. Manuel Osvaldo Machado Rivero<sup>1</sup> mosvaldo@uclv.edu.cu  
Dr.C. Grizly Meneses Placeres<sup>1</sup> grizly@uclv.edu.cu  
Dr.C. Amed Abel Leiva Mederos<sup>1</sup> amed@uclv.edu.cu  
Dr.C. Febe Angel Ciudad Ricardo<sup>2</sup> fciudad@uci.cu  
Marc Goovaerts<sup>3</sup> marc.goovaerts@uhasselt.be  
M.Sc. Nirma Maria Acosta Nuñez<sup>4</sup> nirmacosta@upr.edu.cu  
Dr.C. Ronal Tamayo Cuenca<sup>5</sup> ronaltc@uho.edu.cu  
M.Sc. Leandro Tabares Martin<sup>2</sup> ltmartin@uci.cu  
M.Sc. Erdin Espinosa Gonzalez<sup>4</sup> erdin.espinosa@upr.edu.cu

1-Dirección de Información Científico Técnica de la Universidad Central “Marta Abreu” de Las Villas.  
Cuba.

2-Universidad de las Ciencias Informáticas. Cuba.

3-Universidad de Hasselt. Bélgica.

4-Universidad de Pinar del Rio “Hermanos Sainz”. Cuba

5-Universidad de Holguín “Oscar Lucero Moya”. Cuba

**Resumen:** El objetivo de este trabajo es describir las principales características de la red de repositorios digitales institucionales desarrollada por el proyecto ELINF, los niveles de interoperabilidad entre sus componentes y el desarrollo de un cosechador nacional.

Se realizó un análisis descriptivo que caracteriza el rol del proyecto ELINF en el incremento de la visibilidad académica de cinco universidades cubanas a través del desarrollo de repositorios digitales en esas instituciones; así como el trabajo que se ejecuta para la consolidación de una red nacional. También se describen las herramientas desarrolladas en el marco del proyecto y las líneas de trabajo para garantizar la interoperabilidad entre repositorios y otras plataformas.

El proyecto ELINF, a través de la implementación de repositorios digitales, ha contribuido a potenciar el intercambio de información y conocimientos de varias universidades cubanas, permitido también socializar, preservar e incrementar la visibilidad de la producción científica de tales instituciones que hoy se almacena en sus bibliotecas. Este resultado se alcanza tras el desarrollo de la primera política de acceso abierto dentro del Ministerio de Educación Superior, la génesis de varias herramientas para la gestión de la información, el entrenamiento de recursos humanos para dar soporte al mantenimiento e interoperabilidad de diversas plataformas. En este momento 5 repositorios están operando como resultado de este proyecto, 4 de ellos con visibilidad internacional e indexados por buscadores académicos y otros directorios en la web.

Se presentan las principales características de DarkAIV, una herramienta para la extracción automatizada de metadatos y que facilita la carga de documentos a la plataforma Dspace.

**Palabras Clave:** Repositorios Digitales Institucionales, Acceso Abierto, Educación Superior Cubana, ELINF.

**Eje temático:** Humanidades

**Abstract:** The aim of this work is to describe the main characteristics of the institutional digital repository network developed by the ELINF project, the levels of interoperability between its components and the development of a national harvester.

A descriptive analysis was carried out to describe the role of the ELINF project in increasing the academic visibility of five Cuban universities through the development of digital repositories in these institutions; as well as the work being carried out for the consolidation of a national network. It also describes the tools developed in the framework of the project and the work's line to ensure interoperability between repositories and other platforms.

The ELINF project, through the implementation of digital repositories, has contributed to enhance the exchange of information and knowledge of several Cuban universities, also allowed to socialize, preserve and increase the visibility of the scientific production of such institutions that is now stored in their libraries. This result is achieved after the development of the first open access policy within the Ministry of Higher Education, the genesis of several tools for information management, training of human resources to support the maintenance and interoperability of various platforms. At the moment 5 repositories are operating as a result of this project, 4 of them with international visibility and indexed by academic search engines and other directories on the web.

The main features of DarkAIV are presented, a tool for the automated extraction of metadata and that facilitates the loading of documents to the Dspace platform.

## 1. Introducción

El Movimiento de Acceso Abierto (AA) comenzó en 1993 como una iniciativa del *Open Society Institute*, sin embargo, no fue hasta principios del siglo XXI que este movimiento alcanzó una fuerza real con la Declaración de Budapest firmada por 270 organizaciones y más de tres mil científicos en todo el mundo [1-3]. El AA ha establecido dos vías de despliegue: la ruta dorada y la verde, la última incluye el desarrollo de Repositorios Digitales [4-8].

Los Repositorios Digitales (RD) se han convertido en una herramienta importante para las instituciones académicas y científicas. Las bibliotecas de las universidades han desarrollado RD con el objetivo de facilitar el acceso de los usuarios a sus colecciones digitales y socializar los resultados científicos y académicos (artículos, informes de conferencias, tesis y disertaciones, materiales de enseñanza, etc.) desarrollados por las propias casas de altos estudios [5, 8-10].

El primer repositorio llamado arXiv se desarrolló para el intercambio de *pre-prints* en la Universidad de Los Álamos (USA), siguiendo la propuesta del físico Paul Ginsparg [8]. Hoy en día, más de 3500 repositorios están disponibles en todo el mundo y registrados en el Directorio de Repositorios de Acceso Abierto [11], también se han implementado algunas iniciativas como: Recolecta (Fundación Española para la Ciencia Abierta, 2018, OpenAIRE y la Confederación de Repositorios de Acceso Abierto, con el objetivo de desarrollar estándares, políticas y facilitar la recuperación de información para repositorios [12].

En 2013 cinco importantes universidades cubanas: Universidad Central "Marta Abreu" de Las Villas (UCLV), Universidad de las Ciencias. Informáticas (UCI), Universidad de Pinar del Río "Hermanos Saíz Montes de Oca" (UPR), la Universidad "Oscar Lucero Moya" de Holguín (UHo) y la Universidad "Ignacio Agramonte y Loynaz" de Camagüey (UCm), iniciaron un programa internacional de colaboración patrocinado por el Consejo Interuniversitario de Flandes (VLIR) desde Bélgica [13], que busca crear un entorno virtual de investigación y educación en Cuba utilizando soluciones de software de código abierto [14]. El programa denominado Red de Cooperación Universitaria: "Fortalecimiento del rol de las TIC en las universidades cubanas para el desarrollo de la sociedad" (también conocido como REDTIC) incluye tres proyectos, uno de ellos llamado: *ICT supporting the educational process and the knowledge management in higher education* (ELINF). Este proyecto incluye el desarrollo de RD y su mejora para difundir los resultados científicos y académicos.

La Comisión Nacional Cubana para el desarrollo del Acceso Abierto fue creada en 2009, con el objetivo de impulsar el AA en el país [15]. Los principales logros de este movimiento hasta 2014 se centraron en la implementación y el uso del *Open Journal System* como plataforma para la gestión de revistas científicas cubanas, especialmente en el área de la salud. La expansión de los RD en Cuba muestra un lento crecimiento entre el año 2009 y 2014, que se caracterizó por el desarrollo del repositorio de tesis doctorales de INFOMED y algunas iniciativas con visibilidad internacional como: EcuRed, el RD del Instituto de Geografía Tropical y *Scriptorium* [RD de la Universidad de La Habana] [16].

## 2. Metodología

Análisis descriptivo que caracteriza el rol del proyecto ELINF como ente potenciador de la visibilidad académica de cinco universidades cubanas a través de RD, así como su

desarrollo futuro hasta la consolidación de una red nacional de repositorios. También describe las herramientas desarrolladas en el marco del proyecto y las líneas de trabajo para garantizar la interoperabilidad entre los RD y otras plataformas.

### **3. El rol del proyecto ELINF en el desarrollo de repositorios digitales en las universidades cubanas**

Las cinco universidades miembros del programa REDTIC trabajan como un pequeño consorcio, con un equipo de desarrollo para generar y mejorar herramientas, así como con un comité de expertos para evaluar y aprobar políticas y otras mejoras. Dspace fue la plataforma escogida para crear los RD, este software de código abierto es uno de los más utilizados en todo el mundo para este propósito [17-19].

A principios de 2014, el proyecto ELINF realizó la primera acción con el objetivo de desarrollar una red de repositorios en las universidades cubanas. Cinco bibliotecarios (uno de cada universidad miembro), recibieron una capacitación por parte de varios especialistas en la Universidad de Hasselt (Bélgica) sobre gestión de información, repositorios institucionales y acceso abierto. Al mismo tiempo, se creó un equipo en la UCI integrado por desarrolladores de software, con el objetivo de estimular una comunidad de desarrollo de Dspace en Cuba para apoyar esa plataforma. Al final de ese mismo año, se desarrollaron varios talleres para capacitar más personal en las cinco universidades; y aumentar así sus conocimientos y habilidades sobre Dspace y el AA en general.

Uno de los problemas encontrados al principio por los miembros del proyecto, fue la gran cantidad de tesis digitales y otros materiales almacenados en las bibliotecas, ya que los bibliotecarios debían cargarlos manualmente al RD. Para resolver este problema, estudiantes de UCLV desarrollaron una herramienta llamada DarkAIV. Este último es un software destinado a la extracción de metadatos y la publicación automática de estos junto a sus respectivos documentos en Dspace. Esta herramienta permite el manejo de grandes colecciones de documentos para la creación de bibliotecas digitales, combina la extracción automática utilizando diferentes tecnologías con la revisión manual de metadatos, y proporciona una evaluación de estos últimos según su integridad [14, 20].

El comité de expertos de ELINF propuso un procedimiento común para el desarrollo de los RD, posteriormente la directiva de cada universidad aprobó una política local para el AA y el desarrollo de RD siguiendo la guía proporcionada por ELINF. Como parte de la política, se aprobó un conjunto de metadatos para cada tipo de documento [16], utilizando el *Dublin Core* extendido como base esencial, que permite cumplir con los estándares internacionales establecidos para la descripción, así como la interoperabilidad con otros sistemas que utilizan el protocolo OAI-PMH [21-23]. El conjunto de metadatos se definió de acuerdo con las pautas de exportación de metadatos lanzado por OpenAIRE [24].

Al mismo tiempo que ELINF aumentaba el nivel profesional de bibliotecarios y personal de informática en las cinco universidades sobre AA y RD, REDTIC mejoró la infraestructura de conectividad e intranet de esas instituciones. Las nuevas capacidades instaladas permiten un mayor intercambio de datos, el uso de nuevas plataformas, la

interoperabilidad entre universidades y la visibilidad internacional de servicios institucionales como los RD y otros diseñados para socializar información y conocimiento.

Con DSpace, ELINF se enfrenta a dos desafíos principales para resolver el problema de la calidad e interoperabilidad de los metadatos con otras plataformas: la granularidad de los metadatos y la implementación de sistemas de autoridad. Por lo tanto, se creó un grupo para desarrollar EsFácil como módulo específico de envío, que incluye funcionalidades de extracción automática de metadatos y la gestión de sistemas de autoridad [14].

### 3.1 Repositorios desarrollados por ELINF

Todas las universidades de la red están utilizando la versión 5.7 o 6 de Dspace, pero personalizados a una versión VLIREC de formato Cuba. Están disponibles en internet a través de la red universitaria nacional cubana [14]. En este momento, cuatro de los cinco repositorios de ELINF tienen visibilidad internacional.

#### **Dspace @ UCLV: repositorio digital en la Universidad Central “Marta Abreu” de Las Villas**

Dspace@UCLV (disponible en: <http://dspace.uclv.edu.cu>) fue el primer repositorio implementado por la UCLV y también el primero de ELINF con visibilidad internacional. Fue diseñado y configurado para almacenar, preservar y diseminar la producción científica realizada en la UCLV.

Dspace@UCLV (figura 1) cuenta con 67 comunidades, 3 de ellas están diseñadas para almacenar las tesis de licenciatura o ingenierías, maestrías y doctorados; mientras que las restantes son para la producción científica de más de 50 departamentos académicos y centros de investigación existentes en la UCLV. Este RD fue indexado por Google Académico en 2015 y también registrado en OpenDOAR y OpenROAR. El *Ranking Web of Repositories* (RWR) ha incluido a Dspace@UCLV en sus reportes desde 2016, evidenciando como este repositorio se convirtió en el RD más importante para la producción científica y académica en Cuba. En la edición del RWR publicada en enero de 2019 Dspace@UCLV mantiene su liderazgo y mejoró su posición al puesto 559 [25].



DSpace@UCLV es el Repositorio Institucional de la Universidad Central "Marta Abreu" de Las Villas (UCLV). Es un servicio que permite el libre acceso a un gran número de documentos científicos (tesis, artículos, libros, monografías, ponencias, registros y patentes) producidos por la UCLV. Con este repositorio se pretende contribuir a la conservación del patrimonio científico institucional y a su visibilidad nacional e internacional. El volumen de documentos depositados en DSpace@UCLV crece constantemente gracias a las contribuciones de los autores y la labor de los bibliotecarios.

**Colecciones destacadas**

- Tesis de Pregrado - UCLV
- Tesis de Pregrado - UCP "Félix Varela" (Hasta 2015)
- Tesis de Pregrado - UCCFD "Manuel Fajardo" (Hasta 2015)

**Comunidades**

Elija una comunidad para listar sus colecciones

- Centro de Bioactivos Químicos (CBQ) [5]
- Centro de Estudio de Dirección Empresarial (CEDE) [1]
- Centro de Estudio de Electroenergética (CEE) [13]

**Búsquedas**

**LISTAR**

- Todo el Repositorio
- Comunidades & Colecciones
- Por fecha de publicación
- Autores
- Títulos
- Materias

**MI CUENTA**

- Acceder
- Registro

**ENLACES EXTERNOS**

- Catálogo Colectivo (OPAC)
- Biblioteca Virtual

Figura 1. Página principal de Dspace@UCLV. Fuente: <http://dspace.uclv.edu.cu>

En este momento, Dspace@UCLV almacena más de 10 000 documentos, una cantidad que aumenta año tras año, porque es la universidad más interdisciplinar de Cuba. En 2017, la implementación de este RD ganó el premio del Ministerio de Ciencia Tecnología y Medio Ambiente de Cuba en la provincia de Villa Clara. Durante más de 3 años de funcionamiento, este servicio se ha convertido en el más visitado de la UCLV, lo que contribuye a mejorar la visibilidad y la posición de esta institución en los rankings internacionales.

#### **REDIUC: repositorio de la Universidad "Ignacio Agramonte y Loynaz" de Camagüey**

REDIUC (figura 2) fue el nombre elegido para el RD en la UCM (disponible en: <http://rediuc.reduc.edu.cu>). Al igual que los otros RD de ELINF, REDIUC almacena la producción académica y científica de esta institución. Está indexado por Google Académico y también registrado en OpenDOAR y OpenROAR. REDIUC fue incluido en el RWR aunque debe realizar determinadas mejoras y una mayor difusión de sus potencialidades al mundo para mejorar posiciones.



Figura 2. Página principal de REDIUC. Fuente: <http://rediuc.reduc.edu.cu>

### **ALMA: repositorio digital en la Universidad "Hermanos Saíz Montes de Oca" de Rinar del Río**

El RD de esta universidad se denominó ALMA (disponible en: <http://rc.upr.edu.cu>), en este momento almacena más de 1900 documentos. Al igual que Dspace@UCLV, sus comunidades y colecciones fueron diseñadas de acuerdo con la estructura académica y científica de la universidad. ALMA (figura 3) también está indexado por Google Académico y registrado en OpenDOAR y OpenROAR. ALMA estuvo incluido en las ediciones del RWR correspondientes a 2017 y 2018.

### **IRD: repositorio en Universidad de las Ciencias Informáticas**

IRD no tiene aún visibilidad internacional, el proyecto ELINF está planificando su registro e indexación internacional a fines de 2019. Además de la producción científica de la Universidad de las Ciencias Informáticas (en este momento almacena más de 7500 documentos), IRD incluye el registro de software y otros productos informáticos.

**Alma**

**Especialidades**  
Colecciona las tesis correspondientes a las especialidades de la Universidad de Pinar del Río.

**Tesis Doctorales**  
Colecciona las tesis correspondientes a los Programas de Doctorado de la Universidad de Pinar del Río y otras defendidas en programas externos a la institución

**Otros Documentos Científicos**  
Colecciona artículos científicos, libros y capítulos de libros, ponencias presentadas en eventos por docentes e investigadores de la Universidad

**Tesis de Maestría**  
Colecciona las tesis correspondientes a los Programas de Maestría de la Universidad de Pinar del Río y otras defendidas en programas externos a la institución

**Navegar por**

Autor	Título	Fecha de lanzamiento
Díaz Pérez, Mailydlyn (31)	Actividades físico-recreativas (182)	2010 - 2019 (1371)
Ferragut Rodríguez, Olga (22)	Nuevas Tecnologías para la Educación (91)	2000 - 2009 (646)
Machin Torres, Belisa (22)	Dirección (88)	1991 - 1999 (16)

Figura 3. Página principal de ALMA. Fuente: <http://rc.upr.edu.cu>

**REPOSITORIO DIGITAL**

**Bienvenido al Repositorio Institucional de la Universidad de las Ciencias Informáticas**  
El Repositorio Institucional se implementa con el objetivo de atesorar, divulgar y conservar la memoria científica de la UCI. Permite incrementar la visibilidad, y el prestigio de la Institución, así como el impacto y el reconocimiento de la producción científica de sus investigadores.

**Comunidades**  
Elija una comunidad para visualizar sus colecciones.

- Facultad 1
- Facultad 2
- Facultad 3
- Facultad 4
- Facultad de Ciencias y Tecnologías Computacionales
- Otras Áreas

**Descubrir por:**

Autor	Temática	Fecha de lanzamiento
Febles Estrada, Ailyn (32)	INFORMATICA (3585)	2010 - 2017 (4824)
Carrasco Velar, Ramón (31)	DESARROLLO DE SOFTWARE (3344)	2001 - 2009 (2746)
Antelo Collado, Aurelio (29)	COMPUTACION (1938)	
Ciudad Ricardo, Febe Angel (28)	INGENIERIA DE SOFTWARE (1865)	
Díaz Vera, Julio César (28)	PROGRAMACION (1281)	
Fernández Pérez, Yamilis (28)	SISTEMAS DE GESTION (1104)	

Figura 4. Página principal de IRD

### 3.2 Red nacional de repositorios de ELINF

Con todos los repositorios en línea (al menos a nivel nacional), ELINF está desarrollando un cosechador nacional para hacer posible la búsqueda y recuperación de información a través de una interfaz única. Este recolector (en línea el próximo diciembre del año en curso), será alojado en el centro de datos ubicado en la Universidad Central “Marta Abreu” de Las Villas, y hará posible que cada repositorio actúe como un nodo de red real y no como un sistema independiente, generando una red federada de RD, el intercambio de datos e información será posible a través del Protocolo de la Iniciativa de Archivos Abiertos para la Recolección de Metadatos [OAI-PMH] [26].

Uno de los objetivos ELINF con la evolución de los RD es que esta red inicial se convierta en la Red Nacional de Repositorios del Ministerio de Educación Superior de Cuba (MES). Para lograr ese objetivo, ELINF está trabajando con dicho ministerio como su principal beneficiario en el país.

El MES, tras la experiencia de las cinco universidades incluidas en el entorno de consorcio, decidió liderar un proyecto nacional para expandir esta experiencia al resto de universidades cubanas (más de 21). La interoperabilidad de más de 20 repositorios digitales permitirá la génesis y la gestión en 2020 de la mayor colección digital científica disponible en Cuba, ya que cada año las universidades producen más del 50% de los resultados científicos cubanos.

Una vez que el cosechador nacional esté funcionando, ELINF y el MES solicitarán la inscripción de esta red de repositorios en la Confederación de Repositorios de Acceso Abierto, como un paso más para el reconocimiento y la visibilidad del movimiento de acceso abierto en la educación superior cubana.

### 3.3 Interoperabilidad entre repositorios DSpace y otros sistemas

ELINF también comenzó con la implementación de otras herramientas específicas para la gestión educativa, de bibliotecas, información y resultados de investigación, además del Dspace aparecen dentro del proyecto las plataformas: Moodle, ABCD 3.0 y VIVO. El objetivo final es crear una red integrada de investigación, educación e información.

ABCD 3.0 es el sistema de gestión de bibliotecas oficial aprobado por el MES, este sistema fue utilizado por primera vez en Cuba por la Universidad Central “Marta Abreu” de Las Villas, y después de algunos años de uso de este sistema la UCLV propuso su utilización dentro de todas las instituciones universitarias del MES [13]. Desde 2013, ELINF trabajó en el desarrollo de ABCD 3.0, el que finalmente se lanzó en enero de 2017 e instaló en todas las bibliotecas del MES [14].

VIVO es una herramienta para crear los denominados *Current Research Information Systems* (CRIS), un software de código abierto para administrar el alcance más amplio de la información de investigación [14].

Moodle es la plataforma aprobada por el MES para apoyar la educación en todas las universidades y forma parte de la estrategia nacional para los entornos de aprendizaje virtual.

Las diferentes plataformas utilizadas en la red (Moodle para aprendizaje electrónico, ABCD como sistema de bibliotecas, DSpace como repositorio y VIVO como sistema

de información de investigación), tienen funcionalidades específicas y, por lo tanto, son relevantes por sí mismas. Por lo tanto, el proyecto no quiere crear una capa adicional sobre las plataformas existentes, sino que quiere hacer posible el intercambio de datos y metadatos entre ellas [14].

ABCD, DSpace, Moodle y VIVO están utilizando diferentes formatos de metadatos y ontologías, debido a la cultura y las necesidades de las comunidades de bibliotecas, repositorios y educativas, respectivamente. Pero como Erik Duval declaró: 'La interoperabilidad es permitir que la información que se origina en un contexto se use en otro de manera tan altamente automatizada como sea posible' [27]. Un formato de metadatos único para todas las plataformas no es realista, pero claramente existe la necesidad de una conversión compartida para identificar y codificar valores antes de poder intercambiar metadatos.

El grupo de trabajo ELINF para los sistemas de información de investigación evaluó los elementos y los metadatos relevantes para su reutilización en las diferentes plataformas. La Tabla 1 muestra los principales elementos relevantes para la reutilización, en rojo son los que en este momento se pueden definir de forma única.

**Tabla 1.** Metadatos para su reutilización. Fuente: Goovaerts, Ciudad Ricardo (14)

Moodle Course information	DSpace Publications	VIVO Publications	ABCD Books
Teacher	Authors	Authors,	Authors,
Title	Title	Title	Title
Course level	Organization	Organization	Organization
Finality	Keywords -	Keywords -	Keywords -
Abstract	subject	subject	subject
Organization	Source:	Source:	ISBN
Keywords -	- Journal title (issn)	- Journal title (issn)	DOI
subject	ISBN	ISBN	Date
Date (?)	DOI	DOI	Abstract
	Date	Date	Publisher
	Abstract	Abstract	
	Publisher	Publisher	

Los elementos comunes en las diferentes plataformas son los que definen personas, organizaciones y contenido. El problema que enfrenta ELINF ahora mismo es cómo se pueden definir en el formato de metadatos de las diferentes plataformas y qué archivos de autoridad y vocabularios controlados se pueden usar, como se pueden implementar identificadores únicos [14].

Las plataformas tienen que usar el mismo archivo de autoridad para investigadores, organizaciones y académicos. Se debe agregar una ID única a los formatos de metadatos.

Una forma de definir el contenido de manera inequívoca es mediante el uso de vocabularios controlados. Hay diferentes tesauros traducidos a SKOS. ELINF ya usa CCS, AGROVOC, MESH y otros tesauros se pueden agregar. Si algunos otros tesauros no están disponibles en formato SKOS, pueden transformarse utilizando herramientas como VocBench [14].

*EsFácil Authority* es una herramienta desarrollada en el marco del proyecto ELINF,

para gestionar el acceso a diferentes vocabularios y sistemas de autoridad a nivel local. Está completamente implementada basado en el software libre y de código abierto (FOSS), que permite compartir *EsFácil Authority* con la comunidad internacional como FOSS a través de GitHub: <https://github.com/ELINF-Cuba-Network>. Esta herramienta tiene como objetivo crear un ecosistema de aplicaciones que permita la capacidad de control de autoridad para aplicaciones externas, mediante la reutilización de datos estructurados semánticamente compartidos por diferentes instituciones [14].

Después del desarrollo del sistema de autoridad, *EsFácil* debe desarrollarse como un módulo de envío para crear metadatos ricos. Este trabajo aún está en progreso, pero se realizan los componentes principales, la extracción automatizada de metadatos con DarkAIV y el control de autoridad (*EsFácil Authority*). *EsFácil* se concibe como un módulo independiente que ingerirá registros y archivos en repositorios y sistemas de información, en primer lugar, en DSpace. La idea es almacenar metadatos ricos y traducirlos al formato de las aplicaciones de destino, haciendo posible la interoperabilidad.

Un *Enterprise Service Bus* (ESB) es otra forma que los programadores de ELINF están desarrollando para hacer posible la interoperabilidad entre plataformas. ESB es una colección de servicios de *middleware* que proporciona capacidades de integración. Estos servicios de *middleware* se encuentran en el corazón de la arquitectura ESB en la que las aplicaciones colocan mensajes para enrutar y transformar [28].

#### 4. Conclusiones

El proyecto ELINF ha demostrado el logro del trabajo de consorcio entre bibliotecas universitarias cubanas, desarrollando herramientas y políticas para la gestión de la información, aumentando la visibilidad e intercambio de resultados científicos y académicos de cinco instituciones a nivel nacional e internacional. Las experiencias logradas para ELINF se han compartido con el resto de las universidades cubanas y podrían extenderse a otras organizaciones en el país como una forma de mejorar la información y la gestión del conocimiento en Cuba.

El acceso abierto es una opción viable para acceder y compartir recursos de información para países cuyos recursos financieros no permiten el acceso a bases de datos internacionales grandes y costosas.

La red de repositorios de ELINF se ha convertido en el servicio en línea más importante de Cuba para compartir y aumentar la visibilidad de la producción científica y académica de la educación superior cubana. Asimismo, ha estimulado a otras instituciones a seguir esta iniciativa y hacer posible la primera red de repositorios cubanos: la Red Nacional de Repositorios del Ministerio de Educación Superior de Cuba.

#### Agradecimientos:

Programa Internacional: *Network of University Cooperation "Strengthening the role of ICT in Cuban universities for the development of society" (REDTIC)*. Proyecto 3: *ICT supporting the*

*educational process and the knowledge management in higher education (ELINF)*. Patrocinado por el Consejo Interuniversitario Flamenco (VLIR) de Bélgica.

## Referencias

1. Melero R. Veinte años inmersos en la cronología del acceso abierto a la ciencia. RELIEVE [Internet]. 2014 [consultado 23 may 2018]; 20(2):[aprox. p.]. Disponible en: DOI: 10.7203/relieve.20.2.4300.
2. N. van Leeuwen T, Tatum C, Wouters PF. Exploring Possibilities to Use Bibliometric Data to Monitor Gold Open Access Publishing at the National Level. JOURNAL OF THE ASSOCIATION FOR INFORMATION SCIENCE AND TECHNOLOGY. 2018;69(9):1161-73.
3. Ruiz Pérez S, Delgado López-Cózar E. Spanish researchers' opinions, attitudes and practices towards open access publishing. El profesional de la información. 2017;26(4):722-73.
4. Aleixandre Benavent A, Ferrer sapena A, Alonso arroyo A, Vidal infer A, Domínguez R, González de Dios J. Comunicación científica (XXVI). Cómo aumentar la difusión y el impacto de los trabajos pediátricos participando en la ciencia abierta. Acta Pediátrica Española. 2015;73(8):203-10.
5. Cano Inclán A, De Dios Arias RA, García García O, Cuesta Rodríguez F. Institutional repositories: current status in the international, Latin American and Cuban environment. Revista Cubana de Información en Ciencias de la Salud [Internet]. 2015 [consultado 18 mayo 2019]; 26(4):[aprox. 9 p.]. Disponible en: [http://scielo.sld.cu/scielo.php?script=sci\\_arttext&pid=S2307-21132015000400002&lang=pt](http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S2307-21132015000400002&lang=pt).
6. Lamani MB, Patil RR, Kumbar BD. Open Access E-books in Social Science: A Case Study of Directory of Open Access Books. Journal of Library & Information Technology. 2018;38(2):141-4.
7. Tang M, Bever JD, Yu FH. Open access increases citations of papers in ecology. Ecosphere [Internet]. 2017 [consultado 23 marzo 2019]; 8(7):[aprox. 7 p.]. Disponible en: e01887.10.1002/ecs2.1887.
8. Torres Salinas D, Robinson García N, Aguillo IF. Bibliometric and benchmark analysis of gold open access in Spain: big output and little impact. El Profesional de la Información. 2016; 25(1):17-24.
9. Bijan Kumar Roy, Parthasarathi Mukhopadhyay, Subal Chandra Biswas, Rajesh Das. Developing Open Access Institutional Digital Repository Using Open Source Software: A Step by Step Guide. Journal of Library & Information Science [Internet]. 2017 [consultado 2 mayo 2019]; 7(2):[aprox. 6 p.]. Disponible en: <https://search.proquest.com/openview/544ce270ac6bc5072692cdd035c083a1/1?pq-origsite=gscholar&cbl=1246355>.
10. Machado Rivero MO, Rodríguez Hidalgo RC, Rivero Cañizarez DJ, Leiva Mederos AA, Lorenzo Rodríguez MR. Dspace@UCLV: Repositorio Digital Institucional para la producción científica de la Universidad Central "Marta Abreu" de las Villas. Congreso Universidad [Internet]. 2016 [consultado 2 may 2019]; 5(5):[aprox. 10 p.]. Disponible en: <http://www.congresouniversidad.cu/revista/index.php/congresouniversidad/article/view/1248>.
11. University of Nottingham. OpenDOAR: Directory of Open Access Repositories [Internet]. Nottingham, UK: University of Nottingham; 2018 [consultado 2 mayo 2019]. Disponible en: <http://www.opendoar.org/index.html>.

12. Confederation of Open Access Repositories. Confederation of Open Access Repositories: towards a global knowledge commons [Internet]. Germany: COAR; 2018 [consultado 2 mayo 2019]. Disponible en: <https://www.coar-repositories.org/>.
13. Machado Rivero MO, Alvarez LC, Ciudad Ricardo FA, Goovaerts M, Leiva Mederos AA, Meneses Placeres G, et al., editors. Sharing scientific and academic outputs from Cuban universities through a network of digital libraries. IFLA WLIC 2017 - Libraries Solidarity Society Paper presented in Session 219 - Acquisition and Collection Development; 2017 August; Wrocław, Poland. Wrocław: IFLA. Disponible en: <http://library.ifla.org/id/eprint/1774>
14. Goovaerts M, Ciudad Ricardo FA, Alvarez Fernández LC, Tabares Martín L, Leiva Mederos AA, Machado Rivero MO, editors. Metadata and Authority Systems for VLIREC Cuba. International Conference INFO 2018 III International Workshop of Semantic Web; 2016 12 - 16 march 2018; Havana: CEUR-WS. Disponible en: <http://ceur-ws.org/Vol-2096/paper4.pdf>
15. Casate Fernández R, Senso J. The Landscape of Open Access Journals in Cuba: the strategy and model for its development. In: Rudasill L, Dorta Duque M, editors. Open Access and Digital Libraries Social Sciences Libraries in Action. Boston2013. p. 89-139.
16. Machado Rivero MO, Didiosky Benitez Erice D, Leiva Mederos AA, Rodríguez Hidalgo RC, Fimia León Y, editors. Dspace@UCLV: Institutional Digital Repository for scientific production at Universidad Central "Marta Abreu" de Las Villas. IFLA WLIC 2016 - Connections Collaboration Community in Session S06 - Satellite Meeting: Document Delivery and Resource Sharing In: Transforming resource sharing in a networked global environment; 2016 10-12 August 2016; Washington, D.C: IFLA. Disponible en: <http://library.ifla.org/1931/1/S06-2016-machado-rivero-en.pdf>
17. Medina González A. Implementación de un repositorio digital para el entorno local de la Facultad de Comunicación de la Universidad de La Habana. Bibliotecas Anales de Investigación. 2017;13(2):202-14.
18. Puerta Díaz M, Martí Lahera Y, Bisset Alvarez E. Federación de repositorios universitarios: realidad y retos en Cuba. In: Universidade Estadual de Londrina, editor. XIX ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – ENANCIB 2018; 22 a 26 de outubro de 2018; Londrina – PR, Brasil: Universidade Estadual de Londrina; 2018.
19. Pyrounakis G, Nikolaidou M, Hatzopoulos M. Building Digital Collections Using Open Source Digital Repository Software: A Comparative Study. International Journal of Digital Library Systems [Internet]. 2014 [consultado 3 mayo 2019]:[aprox. 9 p.]. Disponible en: <http://www.igi-global.com/article/building-digital-collections-using-open-source-digital-repository-software/105108>.
20. Enriquez Rodríguez Rodríguez FA, Hernández Morales LD. Darkaiv: sistema para la extracción y publicación automática de metadatos de publicaciones científicas [Tesis de pregrado]. Santa Clara: Universidad Central "Marta Abreu" de Las Villas; 2016. Disponible en: <http://dspace.uclv.edu.cu:8089/handle/123456789/6703>
21. Anibaldi S, Yaques Y, Celli F, Stellto A, Keizer J. Migrating bibliographic datasets to the Semantic Web: The AGRIS case. Semantic Web [Internet]. 2015 [consultado 23 may 2018]; 6(2):[aprox. p.]. Disponible en: <http://content.iospress.com/articles/semantic-web/sw128>.
22. Fimia Leon Y, Benitez Erice D, Machado Rivero M, Leiva Mederos AA, Álvarez Fernández LC, Martínez Calvet MI, et al. Propuesta de directrices y metadatos para los repositorios institucionales de las universidades de la Red TIC. Santa Clara, Cuba: Samuel Feijóo; 2017.
23. Gregory L, Williams S. On Being a Hub: Some Details behind Providing Metadata for the Digital Public Library of America. D-Lib Magazine [Internet]. 2014 [consultado 2

- mayo 2019]; 20(7/8):[aprox. 7 p.]. Disponible en: <http://www.dlib.org/dlib/july14/gregory/07gregory.html>.
24. Artini M, Atzori C, Bardi A, La Bruzzo S, Manghi P, Mannocci A. The OpenAIRE Literature Broker Service for Institutional Repositories. D-Lib Magazine [Internet]. 2015 [consultado 2 mayo 2019]; 21(11/12):[aprox. 6 p.]. Disponible en: <http://www.dlib.org/dlib/november15/artini/11artini.html>.
  25. Cybermetrics Lab. TRANSPARENT RANKING: Institutional Repositories by Google Scholar (January 2019) [Internet]. 2019 [consultado 23 march 2019]. Disponible en: <http://repositories.webometrics.info/en/node/32>.
  26. Open Archives Initiative Object Exchange and Reuse [Internet]. Nueva York: Open Archives; 2016 [consultado 23 may 2018]. Disponible en: [www.openarchives.org/ore](http://www.openarchives.org/ore).
  27. Duval E, Hodgins W, Sutton S, Weibel SL. Metadata Principles and Practicalities. D-Lib Magazine [Internet]. 2002 [consultado 23 may 2018]; 8(4):[aprox. p.]. Disponible en: <https://doi.org/10.1045/>.
  28. Binildas CA. Service Oriented Java Business Integration. Enterprise Service Bus integration solutions for Java developers. Birmingham: Packt Publishing; 2008.

## Repositorio Institucional del Centro de Ciencias de la Atmósfera, UNAM

Agustín Fernández Eguiarte, Carolina Castelán Hernández y Ángela Galván Mejía

Centro de Ciencias de la Atmósfera  
Universidad Nacional Autónoma de México  
Ciudad Universitaria, CdMx, 04510, México  
[agustin@unam.mx](mailto:agustin@unam.mx), [ccastelanh@atmosfera.unam.mx](mailto:ccastelanh@atmosfera.unam.mx), [angame4@gmail.com](mailto:angame4@gmail.com)

**Resumen.** El Repositorio Institucional del Centro de Ciencias de la Atmósfera de la Universidad Nacional Autónoma de México facilita la integración, manejo, distribución e interoperabilidad de Datos Abiertos científicos geospaciales generados en diversos proyectos de investigación de la dependencia, incorporando además información especializada sobre las diversas temáticas de investigación que se desarrollan en el Centro de Ciencias de la Atmósfera de la UNAM.

El Repositorio es un desarrollo tecnológico interactivo, escalable y actualizable en Acceso Abierto, interoperable con el Repositorio Nacional de Ciencia Abierta de México y con el Repositorio del Programa INSPIRE (Infrastructure for spatial information in Europe). Su objetivo es proporcionar a la sociedad y a los tomadores de decisiones tanto públicos como privados, datos e información para una mejor gestión de los grandes problemas nacionales vinculados con los fenómenos hidrometeorológicos, climáticos y de cambio climático que afectan al país.

Se trata de una nueva alternativa de desarrollo tecnológico que sirve como punto de partida para la generación de nuevo conocimiento científico y que es de utilidad en todos los niveles educativos porque mediante sus metadatos describe y visualiza en forma sencilla y amigable los datos y la información que contiene, lo que permite avanzar en la apropiación social del conocimiento, es decir, en la comunicación entre los científicos y la sociedad en su conjunto para contribuir en la generación de políticas públicas conjuntas que incidan, entre otros beneficios, en una mejor gestión integral del riesgo y en la cultura de la prevención ante los impactos de desastres asociados a fenómenos hidrometeorológicos, climáticos y de cambio climático.

El Repositorio es un desarrollo tecnológico innovador porque además de describir, visualizar y poner a disposición de los usuarios datos científicos georreferenciados mediante metadatos geospaciales, permite continuar integrando datos e información de las diversas temáticas de las ciencias atmosféricas y ambientales que abordan especialistas de la dependencia. Debido a que su desarrollo se fundamentó en sistemas de código abierto puede ser replicado en otras instituciones que generan grandes volúmenes de datos e información geográfica y que requieren integrarla mediante estándares internacionales de metadatos geográficos.

**Palabras Clave:** Datos geospaciales, metadatos, interoperabilidad, visualización científica

**Eje temático:** Medio Ambiente

## 1. Problemática y contexto en que se propuso el proyecto

El Consejo Nacional de Ciencia y Tecnología (CONACYT) de México [1], emitió la Convocatoria 2015 para desarrollar los Repositorios Institucionales de Acceso Abierto a la Información Científica, Tecnológica y de Innovación [2]. En el marco de dicha Convocatoria, el Centro de Ciencias de la Atmósfera (CCA) [3] de la Universidad Nacional Autónoma de México (UNAM) [4] presentó la propuesta denominada “Repositorio Institucional del Centro de Ciencias de la Atmósfera UNAM”, la cual se aprobó y se desarrolló entre agosto de 2017 y marzo de 2018.

Con fundamento en la Ley de Ciencia y Tecnología la Convocatoria establece las bases para promover el desarrollo, la vinculación y diseminación de la investigación científica que se derive de las actividades de investigación básica y aplicada, el desarrollo tecnológico de calidad y la innovación, asociados a la actualización y mejoramiento de la calidad de la educación y la expansión de las fronteras del conocimiento apoyándose en las nuevas tecnologías de la información y, en su caso, mediante el uso de plataformas de acceso abierto, así como convertir a la ciencia, la tecnología y la innovación en elementos fundamentales de la cultura general de la sociedad.

Otro de sus objetivos es promover la creación de programas y espacios públicos virtuales para la apropiación social de la ciencia, la tecnología y la innovación; así como el establecimiento de mecanismos para que la sociedad tenga acceso abierto al conocimiento generado con financiamiento público.

La propia Ley de Ciencia y Tecnología, en su Capítulo X, Artículo 64 [5], establece que: “las instituciones de educación superior y Centros de Investigación podrán constituir Repositorios por disciplinas científicas y tecnológicas u otros que se determinen, a fin de diseminar la información científica y tecnológica que se derive de sus productos educativos y académicos, y en general de todo tipo de investigaciones que realicen, cualquiera que sea su presentación, de acuerdo con criterios de calidad y estándares técnicos que emita el CONACYT. Dichos Repositorios podrán establecerse a nivel de las instituciones y centros de investigación o mediante la creación de redes o asociaciones con otras instituciones, por disciplinas, por regiones u otros”.

## 2. Descripción de la solución tecnológica implementada

### 2.1 GeoNetwork Opensource

El Repositorio se desarrolló con base en la tecnología del proyecto [GeoNetwork Opensource](#) [6] que es una aplicación informática de software libre y código abierto de catalogación para recursos referenciados al espacio geográfico. Es decir, es un catálogo de información orientada a lugares.

GeoNetwork dispone de un mecanismo para la edición de metadatos geográficos, funciones de búsqueda y visualización interactiva de mapas en Internet utilizado actualmente en numerosas iniciativas y proyectos alrededor del planeta. De igual forma

proporciona una interfaz de utilización sencilla y amigable para la búsqueda de datos geoespaciales a través de múltiples catálogos. Los usuarios pueden afinar la búsqueda y rápidamente obtener los registros de su interés.

Las capas geoespaciales e incluso los servicios, mapas o conjuntos de datos no geográficos pueden ser descritos en el catálogo. La descripción de la información se incorpora mediante el editor de metadatos, el cual es compatible con estándares internacionales para recursos geoespaciales y para documentación no espacial. Asimismo, con GeoNetwork es posible crear esquemas de metadatos particulares.

Los protocolos que maneja el sistema GeoNetwork para la interoperabilidad de los datos y la información, son los siguientes: Open Geospatial Consortium Catalog Service for the Web (OGC CSW) [7], Open Archives Initiative – Protocol for Metadata Harvesting (OAI-PMH) [8], OpenSearch [9] y Z39.50 [10]. El protocolo utilizado para la interoperabilidad de los datos y los recursos documentales en el Repositorio Institucional del Centro de Ciencias de la Atmósfera de la UNAM, fue el OAI-PMH conforme a los Lineamientos de la Convocatoria 2015 para desarrollar los Repositorios Institucionales de Acceso Abierto a la Información Científica, Tecnológica y de Innovación.

## **2.2 Metadatos geoespaciales**

Los metadatos geográficos se definen como los datos que describen los datos o información sobre los datos. Conforman un conjunto de información estructurada la cual describe los datos almacenados en sistemas administrativos. Los metadatos pueden proporcionar un breve resumen sobre el contenido, propósito, calidad y ubicación geográfica de los datos, así como información sobre su creación.

Los estándares de metadatos proporcionan a los generadores de datos un conjunto de términos y definiciones documentados en un formato estructurado para describir adecuadamente sus datos, facilitando a los usuarios su aplicación en diversas disciplinas y necesidades específicas.

Los metadatos estructurados en estándares geoespaciales constituyen un soporte eficiente de acceso a los datos mediante un conjunto de elementos y terminologías comunes que permiten búsquedas eficientes y expeditas de los datos requeridos. Los metadatos basados en estándares geográficos describen la consistencia y calidad de la información evitando la pérdida de partes importantes sobre el conocimiento de los datos.

Datos geoespaciales es todo aquel conjunto de datos con una componente geográfica o de ubicación espacial; son usualmente generados por organismos o instituciones especializadas que tienen como fin ponerlos al alcance y al servicio de diversos usuarios y facilitar la perspectiva de la apropiación social del conocimiento, así como que sean de utilidad para analistas, investigadores, programadores, desarrolladores de sistemas de información y tomadores de decisiones tanto públicos como privados.

La catalogación adecuada y estandarizada de los datos geoespaciales facilita a los usuarios una mejor evaluación de los datos para optimizar su producción, almacenamiento y actualización.

Los estándares de metadatos que incorpora el sistema GeoNetwork open source en su versión: 3.0.4.0 son el ISO 19115:2003 de la International Organization for

Standardization (ISO) [11] y el Dublin Core de la Dublin Core Metadata Initiative (DCMI) [12].

El estándar ISO 19115:2003 y sus correspondientes complementos: ISO 19115-1:2014, ISO 19115-2:2019 (previa ISO 19115-2:2009 - Part 2), ISO/TS 19115-3:2016 e ISO/TS 19139-2:2012, es el estándar recomendado por el Digital Curation Centre (DCC) [13] para la descripción de datos e información de las geociencias.

Dublin Core es un modelo de metadatos elaborado y auspiciado por la Dublin Core Metadata Initiative. La International Organization for Standardization lo define en su norma ISO 15836 del año 2009 y la National Information Standards Organization (NISO) en su norma NISO Z39.85-2012. Está diseñado la descripción de documentos en general.

### 3. Aspectos críticos y relevantes a resaltar y detallar

El Repositorio Institucional del Centro de Ciencias de la Atmósfera de la UNAM facilita el manejo, distribución e interoperabilidad de Datos Abiertos geoespaciales integrando también las componentes de artículos científicos y tesis de maestría y doctorado que se generan en la dependencia, extendiendo de este modo su utilidad al conjunto de la sociedad, los tomadores de decisiones tanto públicos como privados y a los sectores educativos y de investigación del país.

El Repositorio es un desarrollo tecnológico innovador porque integra datos científicos geoespaciales con publicaciones científicas, habiendo sido cosechado e incorporado en el Repositorio Nacional de Ciencia Abierta conforme a los lineamientos específicos para Repositorios emitidos por el Consejo Nacional de Ciencia y Tecnología (CONACYT) el 20 de julio de 2017.

El Repositorio es totalmente interactivo, escalable y actualizable en Acceso Abierto interoperable no solo con el Repositorio Nacional de Ciencia Abierta de México y con el Repositorio del Programa internacional europeo para la infraestructura de información espacial, denominado Infrastructure for spatial information in Europe (INSPIRE) [14] para de esta forma proporcionar datos e información para una mejor gestión de los grandes problemas nacionales vinculados con los fenómenos hidrometeorológicos, climáticos y de cambio climático que afectan al país, así como información especializada sobre las diversas temáticas de investigación que se desarrollan en el Centro de Ciencias de la Atmósfera de la UNAM.

Se trata de una nueva alternativa de desarrollo tecnológico que sirve de base para la generación de nuevo conocimiento científico y que es de utilidad en todos los niveles educativos. Mediante sus correspondientes metadatos, describe de manera sencilla y amigable la información que permite avanzar en la apropiación del conocimiento científico por parte de la sociedad en su conjunto, incidiendo en la comunicación entre los científicos y los pobladores de cada una de las más de 300 mil localidades rurales, semiurbanas y urbanas del país.

Para que el proceso de apropiación social del conocimiento sea eficaz deberá partir del interés y de las necesidades de información de los miembros de cada una de las comunidades del país por lo que las componentes de “Datos científicos geoespaciales” del Repositorio, descritas en el apartado 4.1 del presente artículo, contienen en sus

metadatos vínculos a los Servidores de Mapas del Atlas Climático Digital de México y de Escenarios de Cambio Climático.

En los Servidores de Mapas se visualizan interactivamente las bases de datos temáticas y además es posible desplegar y localizar por nombre cada una de las comunidades del país para que los usuarios, partiendo de las localidades en que habitan puedan tener fácil acceso a los metadatos geospaciales que describen y el conocimiento científico actual sobre aspectos hidrometeorológicos, climáticos y de cambio climático a los que se encuentran expuestos e inmersos sus pobladores.

Los servidores de mapas del Atlas Climático Digital de México y de Escenarios de Cambio Climático despliegan cartografía climática y de cambio climático, tanto en formato vectorial (puntos, líneas y polígonos) como con imágenes (mallas) mediante la especificación del estándar Web Map Service (WMS) [15] definida por el Open Geospatial Consortium (OGC) [16].

Los servidores de mapas se desarrollaron en la Unidad de Informática para las Ciencias Atmosféricas y Ambientales ([UNIATMOS](#)) [17] del Centro de Ciencias de la Atmósfera de la UNAM con base en el sistema GeoServer Opensource [18] y permiten la visualización de las bases de datos mediante composiciones cartográficas en diferentes acercamientos combinándolas con diversas capas de referencia. En cualquier posición geográfica se puede solicitar y obtener la información puntual de los datos y de las referencias seleccionadas.

Un aspecto crítico fundamental para el proyecto es contar con apoyo financiero permanente ya que los recursos parciales y temporales del CONACYT, así como la infraestructura y recursos humanos proporcionados por el Centro de Ciencias de la Atmósfera de la UNAM y por la Dirección General de Repositorios Universitarios de la propia UNAM antes denominada Coordinación de Colecciones Universitarias Digitales (CCUD) [19] resultan totalmente insuficientes para su continuidad.

### 3.1 Antecedentes

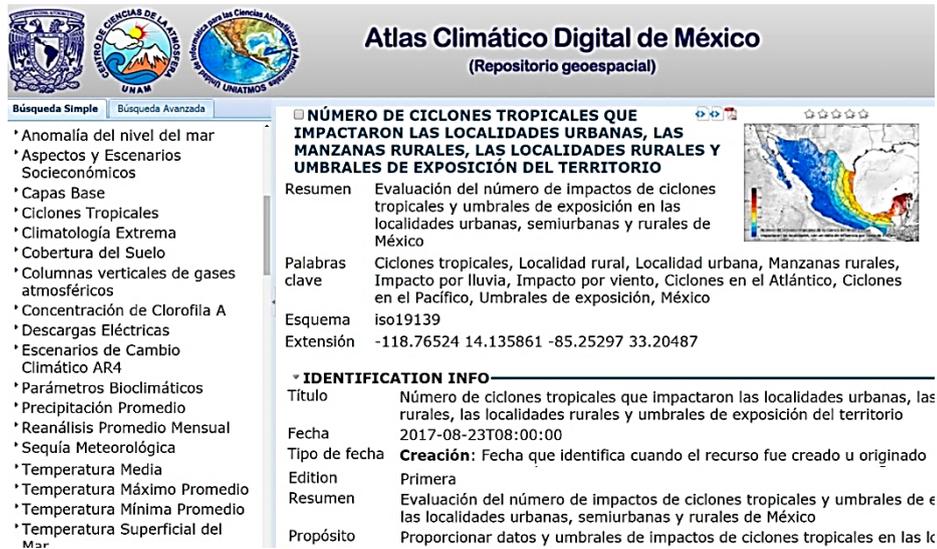
En la UNIATMOS se desarrollaron los siguientes repositorios de datos geospaciales e información ambiental que constituyen los antecedentes del Repositorio Institucional del CCA de la UNAM, así como la totalidad de las bases de datos climáticas y de cambio climático. Las fuentes de las variables climáticas, su control de calidad y la descripción de las metodologías implementadas para la generación de las bases de datos se describen en los metadatos:

- [Repositorio geoespacial del Atlas Climático Digital de México](#) [20]
- [Repositorio geoespacial de escenarios de cambio climático para estudios de impactos, vulnerabilidad y adaptación](#) [21]
- [Repositorio geoespacial de bioclima, cambio climático y ecosistemas en el estado de Tabasco, México](#) [22]

El Repositorio geoespacial del Atlas Climático Digital de México, Figura 1, contiene actualmente 4,626 bases de datos climáticas así como sus correspondientes metadatos que describen las siguientes 21 temáticas o variables climáticas.

- Anomalía del nivel del mar
- Aspectos y escenarios socioeconómicos
- Capas base

- Ciclones Tropicales
- Climatología extrema
- Cobertura del suelo
- Columnas de gases atmosféricos
- Concentración de clorofila a
- Descargas eléctricas
- Escenarios de cambio climático AR4
- Parámetros bioclimáticos
- Precipitación promedio
- Reanálisis promedio mensual
- Sequía meteorológica
- Temperatura máxima promedio
- Temperatura media
- Temperatura mínima promedio
- Temperatura superficial del mar
- Topografía dinámica absoluta
- Velocidad geostrófica
- Viento climatológico en el mar



**Atlas Climático Digital de México**  
(Repositorio geoespacial)

Búsqueda Simple | Búsqueda Avanzada

- \* Anomalía del nivel del mar
- \* Aspectos y Escenarios Socioeconómicos
- \* Capas Base
- \* Ciclones Tropicales
- \* Climatología Extrema
- \* Cobertura del Suelo
- \* Columnas verticales de gases atmosféricos
- \* Concentración de Clorofila A
- \* Descargas Eléctricas
- \* Escenarios de Cambio Climático AR4
- \* Parámetros Bioclimáticos
- \* Precipitación Promedio
- \* Reanálisis Promedio Mensual
- \* Sequía Meteorológica
- \* Temperatura Media
- \* Temperatura Máximo Promedio
- \* Temperatura Mínima Promedio
- \* Temperatura Superficial del Mar

**NÚMERO DE CICLONES TROPICALES QUE IMPACTARON LAS LOCALIDADES URBANAS, LAS MANZANAS RURALES, LAS LOCALIDADES RURALES Y UMBRALES DE EXPOSICIÓN DEL TERRITORIO**

Resumen: Evaluación del número de impactos de ciclones tropicales y umbrales de exposición en las localidades urbanas, semiurbanas y rurales de México

Palabras clave: Ciclones tropicales, Localidad rural, Localidad urbana, Manzanas rurales, Impacto por lluvia, Impacto por viento, Ciclones en el Atlántico, Ciclones en el Pacífico, Umbrales de exposición, México

Esquema: iso19139

Extensión: -118.76524 14.135861 -85.25297 33.20487

**\* IDENTIFICATION INFO**

Título: Número de ciclones tropicales que impactaron las localidades urbanas, las rurales, las localidades rurales y umbrales de exposición del territorio

Fecha: 2017-08-23T08:00:00

Tipo de fecha: **Creación:** Fecha que identifica cuando el recurso fue creado u originado

Edición: Primera

Resumen: Evaluación del número de impactos de ciclones tropicales y umbrales de exposición en las localidades urbanas, semiurbanas y rurales de México

Propósito: Proporcionar datos y umbrales de impactos de ciclones tropicales en las localidades urbanas, semiurbanas y rurales de México

Fig. 1. Portal del [Repositorio geoespacial del Atlas Climático Digital de México](#), en el que se puede acceder, descargar y visualizar los datos y metadatos de 4,626 bases de datos correspondientes a 21 variables climáticas.

**Fig. 2.** Portal del [Repositorio geoespacial de escenarios de cambio climático para estudios de impactos, vulnerabilidad y adaptación](#), en el que se puede acceder, descargar y visualizar los datos y metadatos de 6,984 bases de datos correspondientes a 76 escenarios de cambio climático generados con base en 15 Modelos de Circulación General y tres climatologías de referencia.

**Tabla1.** Modelos de Circulación General contenidos en el Repositorio geoespacial: Actualización de los escenarios de cambio climático para estudios de impactos, vulnerabilidad y adaptación.

Modelo	Institución	País
BCC_CSM1	Beijing Climate Center, China Meteorological Administration	China
CAnESM	Canadian Centre for Climate Modelling and Analysis (CCCma)	Canadá
CNRMCM5	Centre National de Recherches Météorologiques (CNRM-CERFACS)	Francia
CSIRO_Mk3	Commonwealth Scientific and Industrial Research Organisation (CSIRO)	Australia
GFDL_CM3	Geophysical Fluid Dynamics Laboratory (GFDL)	Estados Unidos
GISS_E2_R	NASA Goddard Institute for Space Studies (NASA-GISS)	Estados Unidos
HADGEM2_ES	Met Office Hadley Centre (MOHC)	Reino Unido
INM	Russian Institute for Numerical Mathematics	Rusia
IPSLcm5a_lr	Institut Pierre Simon Laplace (IPSL)	Francia
MIROC_esm	Atmosphere and Ocean Research Institute (The University of Tokyo), National Institute for Environmental Studies, and Japan Agency for Marine-Earth Science and Technology (MIROC)	Japón

MIROC_ESM_CHEM	Atmosphere and Ocean Research Institute (The University of Tokyo), National Institute for Environmental Studies, and Japan Agency for Marine-Earth Science and Technology (MIROC)	Japón
MIROC5	Atmosphere and Ocean Research Institute (The University of Tokyo), National Institute for Environmental Studies, and Japan Agency for Marine-Earth Science and Technology (MIROC)	Japón
MPI_ESM_LR	Max Planck Institute for Meteorology (MPI-M)	Alemani a
MRI_CGCM3	Meteorological Research Institute (MRI)	Japón
NCC_NorESM 1	Bjerknes Centre for Climate Research, Norwegian Meteorological Institute (NCC)	Noruega

El Repositorio geoespacial de escenarios de cambio climático, Figura 2, contiene 6,984 bases de datos de 76 escenarios de cambio climático y sus correspondientes metadatos, algunos de ellos regionalizados para México mediante su referencia a datos climáticos observados en el país.

Los 76 escenarios de cambio climático se desarrollaron a partir de 15 Modelos Generales de Circulación del Panel Intergubernamental sobre Cambio Climático de Naciones Unidas (IPCC, por sus siglas en inglés) [23] considerando los Forzamientos radiativos 4.5, 8.5 (RCP, por sus siglas en inglés) [24] y los Horizontes: Futuro cercano (2015-2039), Futuro medio (2045-2069) y Futuro lejano (2075-2099). Las bases de datos se encuentran disponibles para su descarga en los metadatos correspondientes. Los 15 Modelos Generales de Circulación utilizados, así como la institución y país en donde se generaron se describen en la Tabla 1.

## 4. Resultados obtenidos y su impacto

### 4.1 Repositorio Institucional del Centro de Ciencias de la Atmósfera de la UNAM

El Repositorio Institucional del Centro de Ciencias de la Atmósfera de la Universidad Nacional Autónoma de México (RI) [25] integra la totalidad de las bases de datos del Repositorio geoespacial del Atlas Climático Digital de México y las del Repositorio geoespacial de escenarios de cambio climático, sin embargo, la versión de GeoNetwork utilizada para el Repositorio Institucional es más moderna y corresponde a la versión: 3.0.4.0 por lo que su interface de salida es más amigable y fácil de acceder, mientras que la versión de GeoNetwork empleada en los repositorios antecedentes corresponde a la versión: 2.10.1.

El RI es un repositorio mixto interoperable que incorpora y visualiza bases de datos científicos geoespaciales conjuntamente con artículos científicos y tesis de posgrado que se generan en la dependencia.

Al acceder al repositorio se despliega el portal correspondiente en el que se pueden seleccionar las opciones: “Datos científicos geoespaciales”, “Publicaciones” y “Créditos y cosecha de metadatos”. En “Datos científicos geoespaciales” se pueden

seleccionar “Atlas Climático Digital de México” o “Escenarios de cambio climático”. En la opción “Publicaciones” se puede acceder a “Artículos científicos” o “Tesis”. En la opción “Créditos y cosecha de metadatos” se indican los autores y colaboradores del RI, la dirección para su cosecha, las tecnologías utilizadas en su desarrollo y un agradecimiento al CONACYT por el apoyo al proyecto. Figura 3.

En las opciones: “Atlas Climático Digital de México” y “Escenarios de Cambio Climático”, en sus correspondientes opciones “Buscar”, se pueden realizar búsquedas por: nombre y/o apellido de los autores, cualquier palabra contenida en el metadato, tipo de recurso, categorías, temas, palabras clave, formatos, tipos de representación y escala. Los metadatos geospaciales se encuentran estructurados para las 21 variables climáticas, así como para los 15 Modelos de Circulación General mencionados anteriormente. Los datos correspondientes se encuentran disponibles para su descarga en formatos GeoTIFF y txt y su visualización.

En la opción: “Artículos científicos” y su correspondiente opción “Buscar”, se accede a los metadatos de los artículos científicos y se pueden realizar búsquedas por: nombre y/o apellido de los autores, cualquier palabra contenida en el metadato, palabras clave y años. Los artículos se encuentran disponibles para su visualización y descarga.

En las opciones “Tesis” y “Buscar”, se accede a los metadatos de las tesis y se pueden realizar búsquedas por nombre y/o apellido del autor o director de la tesis, cualquier palabra contenida en el metadato, palabras clave, años y tipo de recurso que pueden ser: Tesis de maestría y Tesis de doctorado. Las tesis se encuentran disponibles para su visualización y descarga.

Repositorio Institucional  
Centro de Ciencias de la Atmósfera, UNAM

Datos científicos geospaciales - Publicaciones - Créditos y cosecha de metadatos

Volver a la búsqueda Atlas Climático Digital de México Artículos científicos  
Escenarios de Cambio Climático Tesis

Vista predeterminada del metadato  
Vista completa del metadato

Precipitación máxima extrema acumulada mensual (mm) en el periodo 1902-2015 y umbrales de exposición del territorio

Visión de Conjunto

Bases de datos de precipitación máxima extrema acumulada mensual y umbrales de exposición del territorio, desarrolladas a partir de los datos de precipitación máxima acumulada mensual de cada una de las estaciones de la...

**Fig. 3.** Portal del Repositorio Institucional del Centro de Ciencias de la Atmósfera de la UNAM en el que se observan las opciones: “Datos científicos geospaciales”, “Publicaciones” y “Créditos y cosecha de metadatos” y sus componentes: “Atlas Climático Digital de México”, “Escenarios de cambio climático”, “Artículos científicos” y “Tesis”. En la figura se despliega la parte superior del metadato correspondiente a “Precipitación máxima extrema acumulada mensual (mm) en el periodo 1902-2015 y umbrales de exposición del territorio”

El Repositorio se puso en línea en marzo de 2018 y una aproximación a su impacto puede estimarse mediante el contador de accesos a cada una de sus componentes. Para la componente Atlas Climático Digital de México se han contabilizado 6,941 accesos al día 30 de julio de 2019, para Escenarios de cambio climático 3,006, para Artículos científicos 6,447 y para la componente Tesis de maestría y doctorado 2,706. El número total al día mencionado es de 19,100 consultas en dieciséis meses.

El Repositorio Institucional del Centro de Ciencias de la Atmósfera de la UNAM constituye un desarrollo tecnológico con un gran potencial de crecimiento que deberá continuar actualizándose y enriqueciéndose con los datos y productos científicos que constantemente producen las investigaciones que se realizan en la institución, así como con las grandes cantidades de bases de datos atmosféricos y ambientales que sistemáticamente se generan mediante las siguientes redes de observación y sistemas de pronóstico:

- Red Universitaria de Observatorios Atmosféricos [26]
- Red del Programa de Estaciones Meteorológicas del Bachillerato Universitario [27]
- Pronósticos meteorológicos [28]
- Pronóstico de oleaje [29]
- Pronóstico de calidad del aire en la CdMx y su zona metropolitana [30]
- Pronóstico de calidad biológica del aire [31]

## **5. Aprendizajes (aciertos y errores)**

La experiencia obtenida en el desarrollo del Repositorio Institucional puede ser aplicable, adaptable y replicable a la gran cantidad de datos geoespaciales que se generan en diversas instituciones académicas y gubernamentales de México e incluso de otros países, que en el mejor de los casos se encuentran disponibles en Acceso Abierto en una gran diversidad de formatos pero sin metadatos geográficos estructurados en estándares internacionales que permitan su interoperabilidad.

Las tecnologías GeoNetwork, ISO 19139 y Geoserver utilizadas en la estructuración del Repositorio Institucional son de igual forma empleadas por el Programa INSPIRE, líder mundial en infraestructuras para generar el futuro de los servicios de datos geoespaciales.

Otra consideración importante es que el desarrollo del Repositorio se basó en sistemas de código abierto lo cual implica que los costos de operación y desarrollo sean considerablemente menores a los que se tienen que erogar cuando se utilizan sistemas de software propietario.

Una condición indispensable para que el proyecto llegue a la población en su conjunto y especialmente la más necesitada y vulnerable, es el acceso eficiente a Internet mediante equipos celulares con tecnología 3G o superiores a fin de que pueda hacerse uso de este y otros servicios educativos mediante las redes. Un problema no resuelto es que las comunidades marginadas del país en las que la desigualdad social se hace presente no cuentan con acceso gratuito y eficiente a estos servicios y tecnologías de la información y la comunicación.

El Repositorio Institucional del Centro de Ciencias de la Atmósfera de la UNAM es el único, entre los repositorios cosechados hasta ahora en el Repositorio Nacional de Ciencia Abierta, que se estructuró con tecnología y estándares de metadatos internacionales geoespaciales como son GeoNetwork e ISO 19115. Los Lineamientos de la Convocatoria 2015 para desarrollar los Repositorios Institucionales de Acceso Abierto a la Información Científica, Tecnológica y de Innovación solo contemplaban al estándar Dublin Core para la cosecha de metadatos en el Repositorio Nacional, por lo que los metadatos geoespaciales originales del Repositorio Institucional hubo necesidad de mapearlos al estándar Dublin Core perdiendo por ello parte de las componentes geográficas del metadato original, sin embargo en el Repositorio Institucional del Centro de Ciencias de Atmósfera de la UNAM se encuentran disponibles los metadatos geográficos originales.

El Repositorio Institucional del Centro de Ciencias de la Atmósfera de la UNAM (primera fase) como se ha mencionado, fue cosechado e incorporado en el Repositorio Nacional de Ciencia Abierta del Gobierno de México, en el que también se han cosechado e incorporado hasta la fecha noventa y dos repositorios de diversas instituciones, sin embargo es importante resaltar que la gran mayoría de dichos repositorios corresponden a recursos documentales de bibliotecas por lo que su desarrollo se ha fundamentado en sistemas y estándares de metadatos diseñados para dicho tipo de recursos, aunque en algunos casos de nuevos repositorios se han incorporado datos geoespaciales en el estándar del metadato internacional denominado DataCite [32] que fue incorporado en lineamientos más recientes.

En la actualidad los metadatos geoespaciales del estándar ISO 19139 del Repositorio Institucional pueden ser cosechados también mediante el estándar DataCite que es un esquema para datos en general, pero que no constituye un metadato geográfico por lo que también limita la estructuración y descripción de datos geoespaciales, aunque con menor severidad que el Dublin Core.

El Repositorio Institucional, como se ha descrito, se desarrolló con base en el sistema GeoNetwork OpenSource; cada una de sus componentes como son, el Atlas Climático Digital de México, Escenarios de Cambio Climático, Artículos científicos y Tesis, constituyen repositorios independientes por lo que las componentes Artículos científicos y Tesis podrían haberse estructurado en DSpace OpenSource [33] que es el sistema más utilizado para catalogar recursos documentales de bibliotecas, sin embargo su interfaz hubiera sido diferente a la de GeoNetwork. Adicionalmente el Repositorio Nacional requería una sola dirección ULR para la cosecha del Repositorio Institucional por lo que se hizo necesario estructurar un quinto repositorio, totalmente transparente para los usuarios, que integra todos los contenidos del Repositorio Institucional. Dicho repositorio se encuentra disponible para su cosecha en la siguiente dirección: <http://ri.atmosfera.unam.mx:8081/geonetwork/srv/spa/oaipmh>

## **6. Impactos no esperados**

Un aspecto no esperado en el que incide y es posible que pueda continuar aportando elementos el desarrollo de Repositorio es en la temática de Gestión Integral de Riesgos tanto a escala nacional, regional, estatal y municipal porque constituye un mecanismo

que facilita la apropiación del conocimiento científico por parte de la sociedad en su conjunto, cobrando relevancia la comunicación de dicho conocimiento entre los pobladores de las diferentes comunidades especialmente las más vulnerables.

De esta forma el Repositorio contribuye en el proceso de transición de un esquema reactivo a uno preventivo ante los impactos y desastres asociados a fenómenos hidrometeorológicos, climáticos y de cambio climático y puede colaborar en la Gestión Integral de Riesgos porque sienta las bases para incorporar muchas más temáticas de riesgo mediante metadatos geospaciales interoperables, integrando, describiendo y visualizando el conocimiento científico en la materia, generado y validado por especialistas por fenómeno natural y antrópico.

Para ello resulta imprescindible que las descripciones de los riesgos asociados a fenómenos naturales y antrópicos contenidas en los metadatos incluyan los aspectos de vulnerabilidad, que sean sencillas, amigables y de fácil comprensión. Asimismo, deben estar vinculadas a mapas interactivos que ubiquen geográficamente dichos fenómenos y sus descripciones. Sin estas consideraciones, los Atlas de Riesgos pueden convertirse en depósitos de información con ninguna o muy limitada función en la Gestión Integral de Riesgos por su difícil interpretación para el conjunto de la sociedad que no está familiarizada en la materia.



**Fig. 4.** Reconocimiento al Repositorio Institucional del Centro de Ciencias de la Atmósfera de la UNAM en el marco de los Premios U-GOB. Proyecto de Datos Abiertos 2019.

El Repositorio Institucional del Centro de Ciencias de la Atmósfera UNAM fue galardonado con el Reconocimiento al Proyecto de Datos Abiertos 2019 en la Categoría Organismo Autónomo Federal en el marco de los [Premios U-GOB 2019 Tecnología en Gobierno](#) [34], entregados el pasado 14 de marzo en el Centro de Innovación y

Desarrollo de la Asociación Nacional de Universidades e Instituciones de Educación Superior (ANUIES). Figura 4.

## Agradecimientos

El proyecto Repositorio Institucional del Centro de Ciencias de la Atmósfera de la UNAM fue financiado parcialmente por el Consejo Nacional de Ciencia y Tecnología de México (CONACYT), Proyecto: 267691, mediante su Convocatoria 2015 para desarrollar los Repositorios Institucionales de Acceso Abierto a la Información Científica, Tecnológica y de Innovación.

De igual forma los autores hacen patente su agradecimiento a la Dirección General de Repositorios Universitarios de la UNAM (anteriormente Coordinación de Colecciones Universitarias Digitales) por su apoyo desde los antecedentes del proyecto.

## Referencias

- 1 Consejo Nacional de Ciencia y Tecnología, <https://www.conacyt.gob.mx/>, Consultado el 9 de mayo de 2019.
- 2 Convocatoria 2015 para desarrollar los Repositorios Institucionales de Acceso Abierto a la Información Científica, Tecnológica y de Innovación, <https://www.conacyt.gob.mx/index.php/el-conacyt/convocatorias-y-resultados-conacyt/convocatorias-direccion-adjunta-de-planeacion-y-evaluacion/convocatoria-2015-repositorios-institucionales-aaict/10723-convocatoria-repositorios-institucionales-2015/file>, Consultado el 9 de mayo de 2019.
- 3 Centro de Ciencias de la Atmósfera, <https://www.atmosfera.unam.mx/>, Consultado el 9 de mayo de 2019.
- 4 Universidad Nacional Autónoma de México, <https://www.unam.mx/>, Consultado el 9 de mayo de 2019.
- 5 Ley de Ciencia y Tecnología, en su Capítulo X, Artículo 64, [http://sep.gob.mx/work/models/sep1/Resource/8dec94b5-f918-45bb-8fd6-adc34a6978da/ley\\_ciencia\\_tecnologia.pdf](http://sep.gob.mx/work/models/sep1/Resource/8dec94b5-f918-45bb-8fd6-adc34a6978da/ley_ciencia_tecnologia.pdf), Consultado el 9 de mayo de 2019.
- 6 GeoNetwork Opensource, <http://geonetwork-opensource.org/>, Consultado el 9 de mayo de 2019.
- 7 Open Geospatial Consortium Catalog Service for the Web, [https://live.osgeo.org/archive/10.0/en/standards/csw\\_overview.html](https://live.osgeo.org/archive/10.0/en/standards/csw_overview.html), Consultado el 9 de mayo de 2019.
- 8 Open Archives Initiative – Protocol for Metadata Harvesting (OAI-PMH), <https://www.openarchives.org/pmh/>, Consultado el 9 de mayo de 2019.
- 9 OpenSearch, <http://www.opensearch.org/Home>, Consultado el 9 de mayo de 2019.
- 10 Z39.50, <https://www.loc.gov/z3950/#about>, Consultado el 9 de mayo de 2019.
- 11 ISO 19115, <http://metadatos.ign.es/normas-familia-iso19100>, Consultado el 13 de mayo de 2019.
- 12 Dublin Core. Dublin Core Metadata Initiative (DCMI), <http://dublincore.org/>, Consultado el 9 de mayo de 2019.
- 13 Digital Curation Centre (DDC). Disciplinary Metadata - Geoscience, <http://www.dcc.ac.uk/resources/metadata-standards/disciplinary/geoscience>, Consultado el 30 de julio de 2019.

- 14 Infrastructure for spatial information in Europe (INSPIRE), <https://inspire.ec.europa.eu/document-tags/network-services>, Consultado el 30 de julio de 2019.
- 15 Web Map Service Open Geospatial Consortium: Web Map Service Implementation Specification (2006), [http://portal.opengeospatial.org/files/?artifact\\_id=14416](http://portal.opengeospatial.org/files/?artifact_id=14416), Consultado el 16 de mayo de 2019.
- 16 Open Geospatial Consortium, <http://www.opengeospatial.org/>, Consultado el 16 de mayo de 2019.
- 17 Unidad de Informática para las Ciencias Atmosféricas y Ambientales (UNIATMOS), Centro de Ciencias de la Atmósfera, UNAM, <http://uniatmos.atmosfera.unam.mx/>, Consultado el 13 de mayo de 2019.
- 18 Geoserver Opensource, <http://geoserver.org/>, Consultado el 13 de mayo de 2019.
- 19 Dirección General de Repositorios Universitarios, UNAM. Antes: Coordinación de Colecciones Universitarias Digitales, UNAM, <http://www.ccu.unam.mx/>, Consultado el 16 de mayo de 2019.
- 20 Repositorio geoespacial del Atlas Climático Digital de México, Centro de Ciencias de la Atmósfera, UNAM, <http://atlasclimatico.unam.mx/geonetwork/>, Consultado el 13 de mayo de 2019.
- 21 Repositorio geoespacial de escenarios de cambio climático para estudios de impactos, vulnerabilidad y adaptación, Centro de Ciencias de la Atmósfera, UNAM, <http://atlasclimatico.unam.mx:8550/geonetwork/>, Consultado el 13 de mayo de 2019.
- 22 Repositorio geoespacial de bioclima, cambio climático y ecosistemas en el estado de Tabasco, México, Centro de Ciencias de la Atmósfera, UNAM, [http://uniatmos.unam.mx:8085/tabasco\\_repositorio/](http://uniatmos.unam.mx:8085/tabasco_repositorio/), Consultado el 13 de mayo de 2019.
- 23 Panel Intergubernamental sobre Cambio Climático de Naciones Unidas, <https://www.ipcc.ch/>, Consultado el 16 de mayo de 2019.
- 24 Representative Concentration Pathways (RCP). Panel Intergubernamental sobre Cambio Climático de Naciones Unidas: Cambio climático. Bases Físicas, Contribución del Grupo de trabajo I al Quinto Informe de Evaluación del Grupo Intergubernamental de Expertos sobre el Cambio Climático: En Glosario, Representative Concentration Pathways (RCP), página 197. Cambridge University Press, Cambridge, Reino Unido y Nueva York, NY, Estados Unidos de América. (2013), [https://www.ipcc.ch/site/assets/uploads/2018/02/AR5\\_WGII\\_glossary\\_ES.pdf](https://www.ipcc.ch/site/assets/uploads/2018/02/AR5_WGII_glossary_ES.pdf), Consultado el 13 de mayo de 2019.
- 25 Repositorio Institucional del Centro de Ciencias de la Atmósfera UNAM, <http://ri.atmosfera.unam.mx>, Consultado el 9 de mayo de 2019.
- 26 Red Universitaria de Observatorios Atmosféricos, Centro de Ciencias de la Atmósfera, UNAM, <https://www.ruoa.unam.mx/>, Consultado el 13 de mayo de 2019.
- 27 Programa de Estaciones Meteorológicas del Bachillerato Universitario, Centro de Ciencias de la Atmósfera, UNAM, <https://www.ruoa.unam.mx/pembu/>, Consultado el 13 de mayo de 2019.
- 28 Pronósticos meteorológicos, Centro de Ciencias de la Atmósfera, UNAM, <http://grupo-ioa.atmosfera.unam.mx/pronosticos/index.php/meteorologia>, Consultado el 13 de mayo de 2019.
- 29 Pronóstico de oleaje, Centro de Ciencias de la Atmósfera, UNAM, <http://grupo-ioa.atmosfera.unam.mx/pronosticos/index.php/oleaje>, Consultado el 13 de mayo de 2019.
- 30 Pronóstico de calidad del aire, Centro de Ciencias de la Atmósfera, UNAM, <https://www.atmosfera.unam.mx/calidad-del-aire-en-mexico>, Consultado el 21 de mayo de 2019.
- 31 Pronóstico de calidad biológica del aire. Red Mexicana de Aerobiología, Centro de Ciencias de la Atmósfera, UNAM, <http://rema.atmosfera.unam.mx/rema/>, Consultado el 13 de mayo de 2019.
- 32 DataCite, <https://datacite.org/>, Consultado el 21 de mayo de 2019.
- 33 DSpace Opensource, <https://duraspace.org/dspace/>, Consultado el 22 de mayo de 2019.

**34** Premios U-GOB 2019 Tecnología en Gobierno,  
<http://132.248.8.242 /u-GOB/>, Consultado el 9 de mayo de 2019.

## **Diseño de interfaces para entornos de interacción performáticos distribuidos de creación sonora y visual**

Mario Humberto Valencia G.  
Manizales Colombia  
Facultad de Artes y Humanidades. Universidad de Caldas

**Resumen.** Este documento se centra en el diseño y en la evaluación de sistemas audiovisuales telemáticos, los cuales permiten a los intérpretes experimentar sonidos, imágenes, expresiones colaborativas y colectivas en ámbitos telemáticos. Basándose en enfoques de interacción humano-computador (HCI) y métodos de usabilidad de tercera ola, en los que se abordan elementos que profundizan más en cómo las personas se relacionan gracias al uso de tecnologías y menos en cómo el usuario cumple sus tareas, se lleva a cabo una serie de estudios de usuarios en diferentes prototipos planteados y desarrollados para tal fin. También se estudian algunas participaciones *performáticas* distribuidas con el objeto de encontrar puntos de inflexión y características útiles que ayuden a las propuestas de prototipos previos así como la construcción de un prototipo impulsado por técnicas de colaboración y evaluación, eje central de la investigación que da pie a este documento, la cual busca diseñar un espacio de comunicación colectiva entre intérpretes telemáticos.

## Introducción

Según MacDonald y Atwood (MacDonald & Atwood, 2013) “la evaluación ha sido un tema dominante en HCI durante décadas, pero está lejos de ser un problema resuelto”. Esto es particularmente evidente en el contexto de los *performances* telemáticos y del desarrollo de interfaces de aplicaciones sonoras e interpretativas en vivo, donde la experiencia del usuario con interfaces lúdicas o creativas es a menudo marcado por una calidad idiosincrásica (Jordà, 2002). Como resultado, las peculiaridades y especificidades en la evaluación de tales interfaces quizá sea un fenómeno inevitable, haciendo que la búsqueda de una solución de “talla única” sea potencialmente inútil. En cambio, se propone que los desarrolladores investiguen la posibilidad de adaptar las técnicas existentes en el diseño de interfaces, un enfoque también defendido por Kiefer, Collins y Fitzpatrick (2008) o diseñar nuevas, si es necesario. La selección de técnicas existentes, como las descritas a lo largo de este documento, puede a su vez estar motivada por varios factores, incluyendo la disponibilidad de las herramientas necesarias, el grado en el que el contexto deseado de una técnica coincide con el examinado y el nivel de modificación requerido para adaptar una técnica de un dominio de aplicación a otro. Adicionalmente, se contempla esta propuesta de estructura metodológica como un claro ejemplo del paradigma de la investigación mixta, ya que se complementan los cuestionarios con discusiones cualitativas abiertas, datos cuantitativos registrados y entrevistas no estructuradas, aumentando así la profundidad de la retroalimentación que pudiéramos obtener de nuestros usuarios. Esto condujo, a su vez, a mejoras en el test y la evaluación con estos permitió reacondicionar y replantear la propuesta de los prototipos.

En los enfoques clásicos del diseño centrado en el usuario, como la usabilidad, se suele poner un fuerte énfasis en el grado en que los usuarios pueden realizar con éxito tareas con un determinado sistema, modelo o prototipo bajo evaluación. Con un número cada vez mayor de disciplinas recurriendo a la investigación de HCI para orientar el diseño de sistemas no solo usables sino también expresivos o simplemente más atrayentes, muchos investigadores se enfrentan a las deficiencias de tal enfoque basado en tareas. Marti y Bannon (2009), por ejemplo, pide un mejor marco para conceptualizar las actividades humanas tanto interpersonales como de comportamiento. De manera similar, Kaye (2007) se pregunta: “¿dónde quedan las estructuras metódicas para el desarrollo de tecnologías planteadas, no para cumplir las tareas, sino para tener experiencias, para expresar la propia identidad, para coquetear y argumentar y vivir?”. A su vez, las insuficiencias del enfoque basado en tareas en el examen de los aspectos de *performance* más allá de la usabilidad, han llevado a la aparición de lo que ahora se conoce como “HCI de tercera onda o de tercer paradigma”, una tendencia descrita por Kiefer, Collins, & Fitzpatrick (2008).

Es así como la evolución de las formas en que la tecnología se utiliza y como la computación se incorpora cada vez más en la vida cotidiana, promueve un enfoque de experiencia en lugar de uno basado en tareas de diseño impulsado por el usuario. En este sentido, estimula lo que describen Fallman y Waterworth (2005) cuando hablan de un enfoque en “experiencias más que en desempeño; diversión y jugabilidad en lugar de tasa de error; y la sociabilidad y las cualidades afectivas más que la capacidad de aprendizaje”. Como resultado, el HCI de tercera ola es particularmente adecuado para el diseño y la evaluación de nuevas interfaces audiovisuales interactivas. Una visión

paralela pero dentro del contexto exclusivo de lo sonoro es presentada por Johnston (2011) y El-Shimy & Cooperstock (2016), quienes desde perspectivas similares postulan que “la evaluación es mejor considerada como un componente de un examen más amplio de la interfaz musical”.

A partir de estos análisis y propuestas, se planea el análisis con la obra telemática, titulada MirroR. Para este análisis se define la aplicación de un test multivariable (Andersen, O'Rourke, Liu, Snider, Lowdermilk, Truong, 2012) y entrevistas semi-estructurada (Wood, 1997), sobre un prototipo de carácter evolutivo (Padhye & Kalia, 2009), haciendo uso de un método orientado hacia el usuario pero aplicando características de experiencia más que de tareas (Doering, Pflieger, Kray, & Schmidt, 2010).

## Procedimientos y diseños

El diseño de entornos de respuesta para el *performance* de interacción distribuida, adoptó un modelo de diseño que hibrida los enfoques descritos. El modelo inicia con una comprensión temprana del usuario objetivo, seguida de ciclos iterativos de pruebas formales, entrevistas semiestructuradas, mejoras en el diseño, ampliación de variables del prototipo, análisis y evaluaciones. La elección de esta metodología particular — impulsada por el usuario— fue motivada en parte por la amplia naturaleza de los objetivos iniciales de investigación.

Si bien se ha establecido la mejora del *performance* de interacción distribuida como el principal objetivo y se han creado directrices para guiar el trabajo, la visión de cualquier funcionalidad concreta en las etapas de diseño era la que podría ayudar a mejorar la experiencia de los artistas distribuidos, el trabajo conjunto con los mismos y permitir la participación en las experiencias del diseño y análisis a todos los participantes, incluidos los músicos. También se tuvo que determinar el tipo específico de intérprete —interactor— hacia el cual los diseños deberían ser adaptados. Es así como se propuso obtener una comprensión de varios tipos de músicos, con un enfoque específico en sus interacciones interpersonales y en sus motivaciones.

## Construcción del test

Después de definir los criterios de evaluación basados en el ensamble, la interacción con los demás, la creatividad y la autoexpresión; y, siguiendo las directrices generales y parámetros impulsores formulados al inicio del proceso de diseño, se encontró que hasta donde sabemos, los métodos estandarizados para evaluar estos factores aún no se han establecido. Herramientas como los cuestionarios de experiencia de juegos abarcan en gran medida preguntas sobre el flujo, la creatividad y la inmersión, esta herramienta demostró ser un elemento adecuado para la evaluación del ensamble y el disfrute debido a la amplitud de los comportamientos examinados (IJsselsteijn, Hoogen, Klimmt, Kort, Lindley, Mathiak, et al., 2008). De hecho, la naturaleza general de las preguntas en los entornos de juego mostraba que era necesaria relativamente poca modificación para adaptarla al contexto musical y audiovisual. Su popularidad y eficiencia motivó el

diseño de los cuestionarios para evaluar puntos de referencia adicionales como la autoexpresión y la creatividad. Para estos cuestionarios también se hizo uso de análisis telemáticos anteriores como los planteados en Telematic Six Perspectives, un paradigmático escrito que propone un cuestionario que, aunque no está estructurado bajo metodologías y técnicas HCI, sí hace uso del amplio conocimiento y experiencia informada de los autores que hacen parte misma de la génesis e implantación de los entornos telemáticos de creación telemática (Oliveros, et al., 2009); de igual forma, fueron usados algunos de los interrogantes y objetivos planteados en los prototipos anteriores, encontrando en ellos importantes variables para ser evaluadas en el prototipo final. Por último se abordaron los parámetros impulsores y las directrices generales de diseño planteadas al inicio del proceso de diseño como guía permanente en el proceso de construcción diseño y evaluación del prototipo.

Es así como se construyó el test para el prototipo MirroR, el cual se encuentra dividido en 4 secciones:

Caracterización del usuario: esta parte del test busca obtener información clara acerca de los participantes a la prueba.

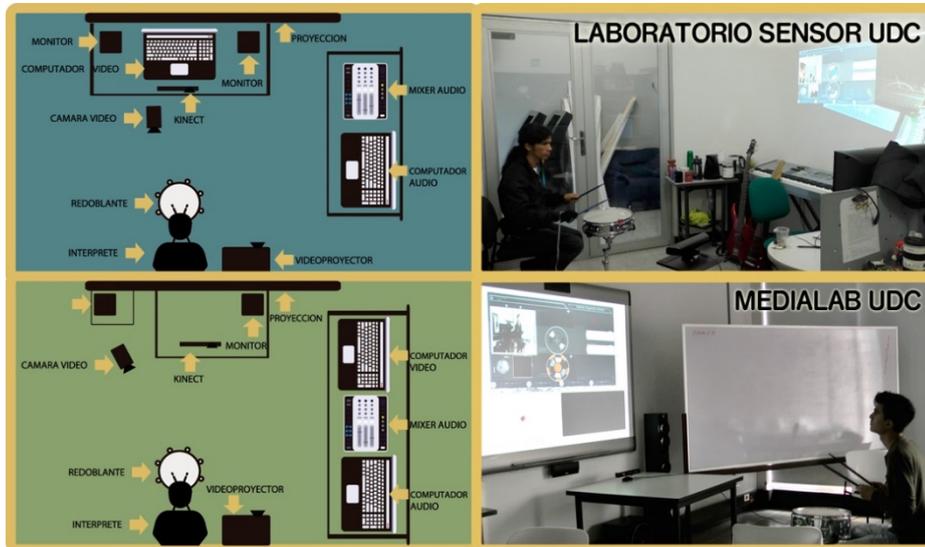
Actuación: este apartado busca conocer cuál es la participación en el *performance* telemático a partir de los criterios descubiertos en el análisis de los prototipos tempranos que permitieron la construcción de roles.

Expectativas: esta sesión busca entender las expectativas de los usuarios a partir de sus perfiles, analizándolos desde sus saberes interpretativos y conocimientos sobre los *performances* telemáticos.

Análisis: los interrogantes planteados en esta sesión buscan dar luces acerca de los procesos y las actuaciones realizadas en la conformación del ensamble y puesta en marcha del *performance* telemático; a continuación, se presentan los ciclos iterativos y pruebas realizadas en cada una de las etapas del *testeo*.

## Ciclos iterativos y pruebas

Para cada sesión de análisis se contó con dos laboratorios de usabilidad en los cuales se ubican cada uno de los intérpretes, un evaluador de experiencia —*tester*—, un asistente de laboratorio, una videocámara, un sistema de audio estéreo, un redoblante, un monitor de video, un videoprojector, un computador, una interfaz de audio, la interfaz MirroR Box y el prototipo MirroR montado (véase la Figura 0.1 Descripción del montaje del laboratorio de análisis multivariable”).



**Figura 0.1** Descripción del montaje del laboratorio de análisis multivariable. Fuente: Elaboración propia a partir de la disposición de los laboratorios de usabilidad, junio 2017.

### Primer ciclo de testeos

Una vez realizados los ajustes pertinentes se inician los ciclos de testeo. Estos ciclos evolutivos parten de la presentación básica de la interfaz y de la adición sistemática de variables a la interfaz; de esta forma, se inicia el primer ciclo de testeos multivariables haciendo uso de dos variables, una de las cuales es fija —el sonido en stream—, a partir de esta condición el sistema propone el siguiente orden de testeos a partir del cambio de variables:

**Tabla 0.1** Definición del primer ciclo de testeos multivariables

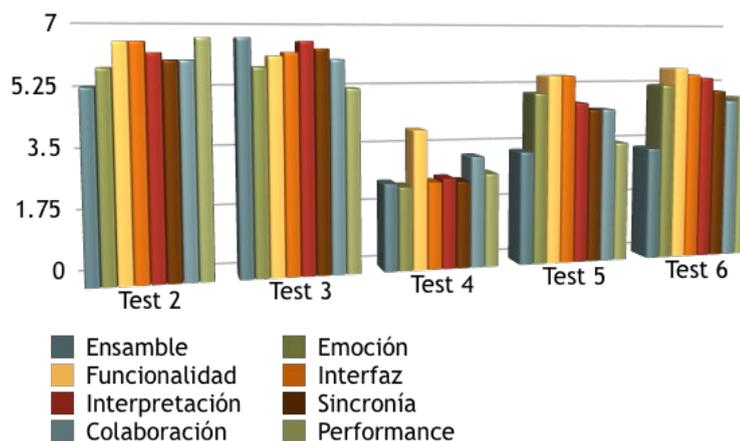
Fuente: Elaboración propia a partir del orden evolutivo de los test planteados, junio, 2017.

Primer Ciclo	Variable 1 - Fija	Variable 2
Test 2	Sonido en <i>stream</i> . JackTrip	Video <i>stream</i> . UG
Test 3	Sonido en <i>stream</i> . JackTrip	Visualización de sensores, giroscopio, acelerómetro y piezoeléctricos. Arduino OSC
Test 4	Sonido en <i>stream</i> . JackTrip	Esqueleto. Kinect OSC
Test 5	Sonido en <i>stream</i> . JackTrip	Visualización 1: nube de puntos
Test 6	Sonido en <i>stream</i> . JackTrip	Visualización 2: superficie de coordenadas

Los datos arrojados al aplicar el primer ciclo de testeos se centran en el uso de las encuestas<sup>1</sup> y de las entrevistas semiestructuradas.<sup>2</sup> Las encuestas son tabuladas y las entrevistas son analizadas y aplicadas a técnica de conteo de ocurrencias, esta cuantificación de actividades se sintetiza en la tabla siguiente:

**Tabla 0.2** Primer ciclo de testeos de cuantificación consolidada. Fuente: Elaboración propia a partir de la cuantificación de categorías analizadas en el primer ciclo de testeos, junio, 2017.

Ciclo	Ensamble	Emoción	Funcionalidad	Interfaz	Interpretación	Sincronía	Colaboración	Performance
Test 2	5,3	5,8	6,5	6,5	6,2	6,0	6,0	6,6
Test 3	6,6	5,8	6,1	6,2	6,5	6,3	6,0	5,2
Test 4	2,5	2,4	4,0	2,5	2,6	2,5	3,2	2,7
Test 5	3,3	5,0	5,5	5,5	4,7	4,5	4,5	3,5
Test 6	3,3	5,2	5,7	5,5	5,4	5,0	4,7	4,8

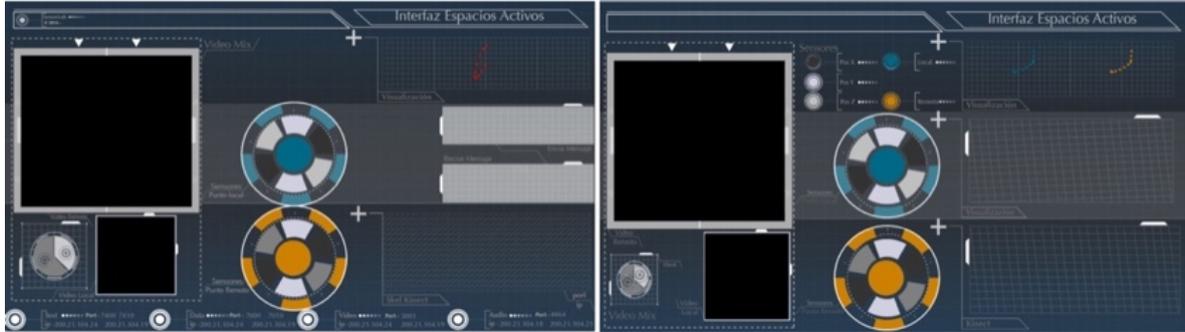


**Figura 0.2** Primer ciclo de testeos. Visualización de la cuantificación. Fuente: Elaboración propia a partir de la cuantificación del primer ciclo de testeos, julio 2017.

Después de este primer ciclo de testeo se observaron algunos inconvenientes en el diseño de la interfaz debido a la saturación de objetos en la interfaz, así como la inoperancia de otros objetos; esto hizo replantear, adaptar y rediseñar algunas características y funcionalidades de la interfaz. También se descubrió —gracias a la sugerencia de los intérpretes y el equipo técnico de testeo— que era necesario el diseño de varias interfaces dependiendo del tipo o característica del usuario, es así como se diseñan tres tipos de interfaces y se rediseña la interfaz de intérprete a partir de las sugerencias de las entrevistas semiestructuradas (figura 0.3).

<sup>1</sup> Los datos obtenidos de las encuestas, a las que se hace referencia, pueden ser visualizado en [https://drive.google.com/open?id=1DH0019x\\_6U5GOamICJFvtNjraaLNw1nB](https://drive.google.com/open?id=1DH0019x_6U5GOamICJFvtNjraaLNw1nB)

<sup>2</sup> Los videos del primer ciclo de testeos y las entrevistas no estructuradas de este ciclo pueden ser consultados en <http://prototipostesis.blogspot.com.co/2017/03/test-1-del-prototipo.html>



**Figura 0.3** Rediseño de interfaz a partir del primer ciclo de testeos. Fuente: Elaboración propia a partir de los cambios realizados en la interfaz MirroR, julio, 2017.

### Segundo ciclo de testeos

Después de realizar las mencionadas correcciones en la interfaz, se programó el segundo ciclo de testeos. En este segundo ciclo se trabajó con tres variables que se agruparon, nuevamente, a partir de los objetos de la interfaz. Para la evaluación de este ciclo se hizo uso de las mismas preguntas del primer ciclo y se realizó un segundo ciclo de entrevistas semiestructuradas. Para este ciclo se planteó la combinación exhaustiva de variables (véase la

Tabla 0.3 Análisis exhaustivo multivariable con tres ítems de análisis), pero con miras a no agotar a los intérpretes, al grupo de testeo y a la poca funcionalidad analítica en el uso de variables exhaustivas (Bernsen & Dybkjær, 2009), se seleccionaron para este ciclo las variables presentadas en la

Tabla 0.4 Ciclo dos: test multivariable con tres variables”.

**Tabla 0.3** Análisis exhaustivo multivariable con tres ítems de análisis

Fuente: Elaboración propia a partir del orden de testeos realizados con tres variables, julio, 2017.

Segundo ciclo	Variable 1: elementos fijos, sonido + partitura	Variable 2	Variable 3
Test 7	Sonido en <i>stream</i> . Partitura	Video <i>stream</i> . UG	Sensores, giroscopio, acelerómetro y piezoeléctricos. Arduino OSC
Test 8	Sonido en <i>stream</i> . Partitura	Video <i>stream</i> . UG	Esqueleto. Kinect
Test 9	Sonido en <i>stream</i> . Partitura	Video <i>stream</i> . UG	Visualización 1: nube de puntos
Test 10	Sonido en <i>stream</i> . Partitura	Video <i>stream</i> . UG	Visualización 2: superficie de coordenadas
Test 11	Sonido en <i>stream</i> . Partitura	Sensores, acelerómetro y Arduino OSC y giroscopio, piezoeléctricos.	Esqueleto. Kinect
Test 12	Sonido en <i>stream</i> . Partitura	Sensores, acelerómetro y Arduino OSC y giroscopio, piezoeléctricos.	Visualización 1: nube de puntos
Test 13	Sonido en <i>stream</i> . Partitura	Sensores, acelerómetro y Arduino OSC y giroscopio, piezoeléctricos.	Visualización 2: superficie de coordenadas
Test 14	Sonido en <i>stream</i> . Partitura	Esqueleto. Kinect	Visualización 1: nube de puntos
Test 15	Sonido en <i>stream</i> . Partitura	Esqueleto. Kinect	Visualización 2: superficie de coordenadas
Test 16	Sonido en <i>stream</i> . Partitura	Visualización 1: nube de puntos	Visualización 2: superficie de coordenadas

**Tabla 0.4** Ciclo dos: test multivariable con tres variables

Fuente: Elaboración propia a partir del orden de testeos realizados con tres variables, julio, 2017.

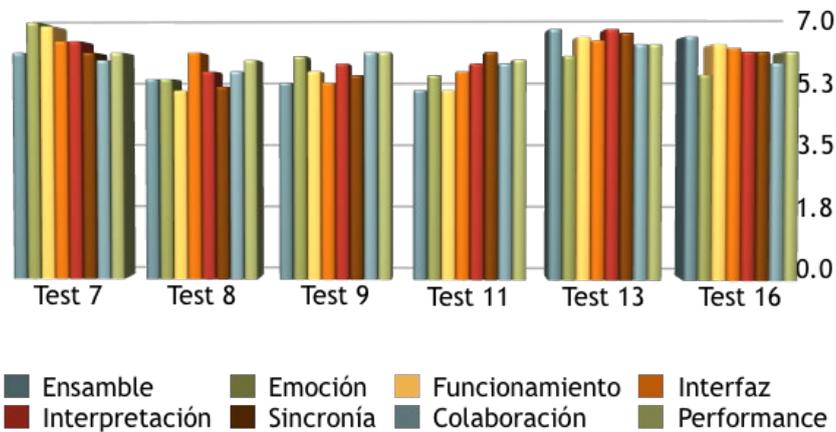
Segundo ciclo	Variable 1: elementos fijos, sonido + partitura	Variable 2	Variable 3
Test 7	Sonido en <i>stream</i> . Partitura	Video <i>stream</i> . UG	Sensores, giroscopio, acelerómetro y piezoeléctricos. Arduino OSC
Test 8	Sonido en <i>stream</i> . Partitura	Video <i>stream</i> . UG	Esqueleto. Kinect
Test 9	Sonido en <i>stream</i> . Partitura	Video <i>stream</i> . UG	Visualización 1: nube de puntos
Test 11	Sonido en <i>stream</i> . Partitura	Sensores, giroscopio, acelerómetro y piezoeléctricos. Arduino OSC	Esqueleto. Kinect
Test 13	Sonido en <i>stream</i> . Partitura	Sensores, giroscopio, acelerómetro y piezoeléctricos. Arduino OSC	Visualización 2: superficie de coordenadas
Test 16	Sonido en <i>stream</i> . Partitura	Visualización 1: nube de puntos	Visualización 2: superficie de coordenadas

De este nuevo ciclo de test se obtuvo un nuevo registro de datos e información que es presentada en la tabla siguiente.

**Tabla 0.5** Segundo ciclo de testeos: cuantificación consolidada

Fuente: Elaboración propia a partir del segundo ciclo de testeos, julio, 2017.

Ciclo 2	Ensamble	Emoción	Funcionamiento	Interfaz	Interpretación	Sincronía	Colaboración	Performance
Test 7	6,0	6,8	6,7	6,3	6,3	6,0	5,8	6,0
Test 8	5,3	5,3	5,0	6,0	5,5	5,1	5,5	5,8
Test 9	5,2	5,9	5,5	5,2	5,7	5,4	6,0	6,0
Test 11	5,0	5,4	5,0	5,5	5,7	6,0	5,7	5,8
Test 13	6,6	5,9	6,4	6,3	6,6	6,5	6,2	6,2
Test 16	6,4	5,4	6,2	6,1	6,0	6,0	5,7	6,0



**Figura 0.4** Segundo ciclo de testeos: visualización de la cuantificación

Fuente: Elaboración propia a partir de la cuantificación del segundo ciclo de testeos, julio, 2017.

A partir de esta información y con este nuevo ciclo de test se pudo analizar que al aumentar la cantidad de variables en la interfaz, las interacciones de los instrumentistas cambiaba con esta y las relaciones entre los intérpretes también. La relación sonido-video-sensores (test7), junto con la partitura, es en la que más cómodos se sintieron, aunque no fueron los mejores valores de ensamble y performance, como se indica más adelante. En las entrevistas semiestructuradas se evidenció que los cambios realizados en la interfaz fueron altamente beneficiosos y no solo para los intérpretes sino también para el equipo técnico; también facilitó, en gran medida, el diseño del escenario.

### Tercer ciclo de testeos

El tercer ciclo de test abordó los siguientes conjuntos de variables.

**Tabla 0.6** Test multivariable ciclo tres, con cuatro variables

Fuente: Elaboración propia a partir del conjunto de variables usado en el tercer ciclo de testeos, julio, 2017.

Tercer Ciclo	Variable 1 - elementos Fijos Sonido + partitura	Variable 2	Variable 3	Variable 4
Test 17	Sonido en <i>stream</i> . Partitura	Video <i>stream</i> . UG	Sensores, giroscopio, acelerómetro y piezoeléctricos. Arduino OSC	Visualización 2: nube de superficie de coordenadas
Test 18	Sonido en <i>stream</i> . Partitura	Video <i>stream</i> . UG	Esqueleto. Kinect	Visualización 1: nube de puntos
Test 19	Sonido en <i>stream</i> . Partitura	Sensores, giroscopio, acelerómetro y piezoeléctricos. Arduino OSC	Visualización 1: nube de puntos	Visualización 2: nube de superficie de coordenadas
Test 20	Sonido en <i>stream</i> . Partitura	Sensores, giroscopio, acelerómetro y piezoeléctricos. Arduino OSC	Esqueleto. Kinect	Visualización 2: nube de superficie de coordenadas

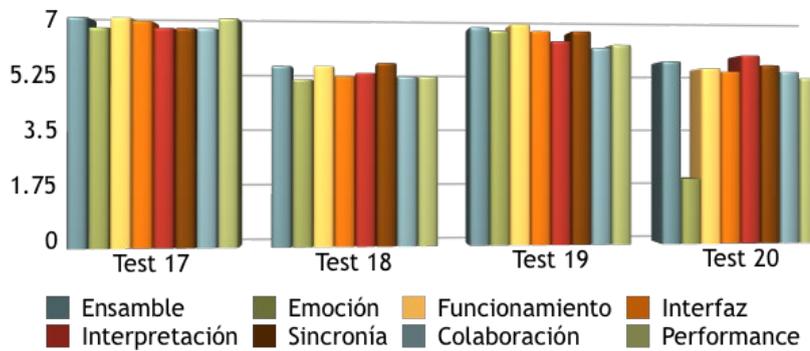
En este nuevo ciclo se simplificaron las preguntas del test en razón a que afloraron algunas categorías de mayor interés en el transcurso de las pruebas anteriores; adicionalmente, algunas preguntas fueron muy repetitivas en el transcurso de los laboratorios ya que las respuestas se consolidaban una y otra vez entre los diferentes test. Por lo anterior, el enfoque de las preguntas se centró en aquellas que brindaran nueva información entre los test y permitieran analizar de forma sintética los cambios observados por los participantes en la evolución de la interfaz. La selección no exhaustiva de las variables también se simplificó a partir de las relaciones entre variables con mejores expectativas, seleccionadas estas a partir de nuevas preguntas, surgidas en el diálogo con los participantes en el montaje de cada test, es de anotar que estos aportes dados por los participantes en los laboratorios, así como los cambios sobre la marcha, nos brindaron gran conocimiento, no solo sobre el prototipo a evaluar sino sobre el método mismo de evaluación impulsada por el usuario.

Fruto de este nuevo ciclo de test y entrevistas semiestructuradas se obtuvieron los datos de la tabla siguiente.

**Tabla 0.7** Tercer ciclo de testeos cuantificación consolidada

Fuente: Elaboración propia a partir de la cuantificación de los datos obtenidos en el tercer ciclo de testeos, julio, 2017.

Ciclo 3	Ensamble	Emoción	Funcionamiento	Interfaz	Interpretación	Sincronía	Colaboración	Performance
Test 17	6,8	6,5	6,8	6,7	6,5	6,5	6,5	6,8
Test 18	5,4	5,0	5,4	5,1	5,2	5,5	5,1	5,1
Test 19	6,6	6,5	6,7	6,5	6,2	6,5	6,0	6,1
Test 20	5,6	2	5,4	5,3	5,8	5,5	5,3	5,1



**Figura 0.5** Tercer ciclo de testeos: visualización de la cuantificación

Fuente: Elaboración propia a partir de la cuantificación de los datos obtenidos en el tercer ciclo de testeos, julio, 2017.

No sin sorpresa encontramos que el test que arrojó mejores características tanto en la interpretación, como en el ensamble y el performance fue el test 17, el cual sintetiza de alguna forma las mejores variables descubiertas en los ciclos anteriores, es de resaltar algunas de las respuestas brindadas por los participantes donde se observa que cada vez más prestan menos atención al stream de video y se concentran más en la representación de los sensores y en las visualizaciones, encontrando incluso que inflexiones específicas de la interfaz funcionaban como diálogos concretos en la interacción entre los instrumentistas, tal vez por esto también encontramos la cercanía en la evaluación entre el test17 y el test19.

#### Cuarto ciclo de testeo

El cuarto y último ciclo de test conjugó el uso de 5 y 6 variables, las variables evaluadas en los test de este ciclo se presentan a continuación.

**Tabla 0.8** Test multivariable: ciclo 4 uso de 5 y 6 variables

Fuente: Elaboración propia a partir de la selección de variables para el cuarto ciclo de tests, julio, 2017.

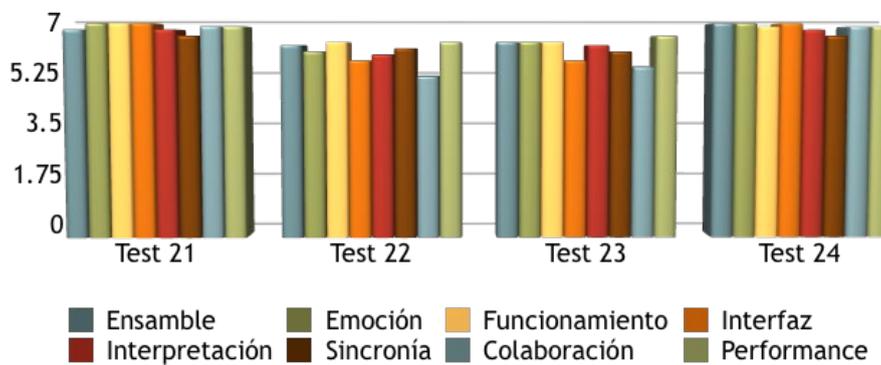
Cuarto Ciclo	Variable 1: elementos fijos sonido + partitura	Variable 2	Variable 3	Variable 4	Variable 5	Variable 6
Test 21	Sonido en stream. Partitura	Video stream. UG	Sensores, giroscopio, acelerómetro piezoeléctricos. Arduino OSC	Visualización 1: nube de puntos	Visualización 2: superficie de coordenadas	-
Test 22	Sonido en stream. Partitura	Video stream. UG	Sensores, giroscopio, acelerómetro piezoeléctricos. Arduino OSC	Visualización 1: nube de puntos	Esqueleto. Kinect	-
Test 23	Sonido en stream. Partitura	Video stream. UG	Visualización 1: nube de puntos	Visualización 2: superficie de coordenadas	Esqueleto. Kinect	-
Test 24	Sonido en stream. Partitura	Video stream. UG	Sensores, giroscopio, acelerómetro piezoeléctricos. Arduino OSC	Visualización 1: nube de puntos	Visualización 2: superficie de coordenadas	Esqueleto. Kinect

En este último ciclo de test se consolidaron algunas de las apreciaciones y hallazgos realizados en los ciclos anteriores, este ciclo fue de gran utilidad ya que en este los participantes dieron apreciaciones generales y propusieron opciones y caminos para desarrollos futuros, la mayor utilidad de este ciclo se dio en los análisis relacionados con los montajes y el performance para el público. Si bien estas preguntas estuvieron siempre en los cuestionarios de los test, fue en este último ciclo donde tanto las preguntas como las entrevistas brindaron más información sobre este campo, es así como las evaluaciones sobre los últimos test reflejan que el ensamble y el montaje fue mejor para el test 24, mientras que tanto el test 24 como el 21 mostraron ser iguales para la sincronía ya que el esqueleto no da valor agregado en esta categoría. En este ciclo se demostró por parte de los intérpretes que el entendimiento de los objetivos de la interfaz y el reconocimiento de la misma permite fomentar en gran medida el ensamble del grupo y que gracias a la presencia de la interfaz como un elemento más en el proceso de construcción y diálogo telemático generó diferentes tipos de actividades y disposiciones que no hubieran existido con un simple modelo de videoconferencia, los participantes también fueron claros en manifestar y evaluar que este cambio en la construcción del performance motivó su participación y frente a la obra, permitiéndoles comprender el entorno telemático como un campo nuevo y diferente que debe ser abordado desde perspectivas diferentes a las tradicionales. Los intérpretes también plantearon la posibilidad de avanzar más en la representación del gesto y en los modelos predictivos de interpretación que de alguna forma admitan trabajar de manera diferente con la latencia, más allá del problema de sincronía y temporalidad.

**Tabla 0.9** Cuarto ciclo de testeos: cuantificación consolidada

Fuente: Elaboración propia a partir de la cuantificación de los datos obtenidos en el cuarto ciclo de testeos, julio, 2017.

Ciclo 4	Ensambl e	Emoción	Funcionamiento	Interfaz	Interpretación	Sincronía	Colaboración	Performance
Test 21	6,7	6,9	6,9	6,9	6,7	6,5	6,8	6,8
Test 22	6,2	6,0	6,3	5,7	5,9	6,1	5,2	6,3
Test 23	6,3	6,3	6,3	5,7	6,2	6,0	5,5	6,5
Test 24	6,9	6,9	6,8	6,9	6,7	6,5	6,8	6,8



**Figura 0.6** Cuarto ciclo de testeos: visualización cuantificación

Fuente: Elaboración propia a partir de la cuantificación de los datos obtenidos en el cuarto ciclo de testeos, julio, 2017.

## Categorización de información, destilando la información

Con miras a aumentar la colaboración entre los intérpretes desplazados y con el objetivo de analizar la construcción y apropiación de los tele-espacios activos, fueron diseñados estos entornos de respuesta y las diferentes estrategias de análisis y evaluación para el performance distribuido. Es así como, mediante el uso de cuatro directrices o características planteadas, a saber: el performance, el ensamble, la interacción y la colaboración, y el tamizado, a partir de los elementos que propusimos para la construcción del espacio activo: interfaz, visualización y emoción, analizamos cómo los intérpretes pueden trabajar, crear y alterar, de forma simple el contexto del performance telemático mismo, al responder interactiva y colectivamente. Este sistema permitió a los intérpretes utilizar la interface y todas las funcionalidades del sistema sin tener que separarse de la tarea principal: su interpretación colectiva. Además, permitió a la investigación analizar la labor de los intérpretes en la construcción del tele-espacio activo, si bien las conclusiones de cada etapa de interrelación entre los intérpretes está sintetizada en los apartados anteriores que se han ido destilando conforme ha avanzado este capítulo, podríamos resumir estos hallazgos y muestras, en la siguiente tabla.

**Tabla 0.10** Hallazgos finales, análisis de la información

Fuente: Elaboración propia a partir de los hallazgos encontrados en el proceso, julio, 2017.

Directrices	Elementos	Previo	Alcanzado
Performance	Interfaz	Metáforas de videoconferencia	Interfaces de representación gestual y predictiva
	Visualización	Stream de video. Masa de movimiento	Comprensión y diseño del escenario activo
	Emoción	Satisfacción al interpretar música compartida	Deseo de participar en nuevos <i>performances</i> telemáticos
	Interpretación	Lectura de partitura	
Ensamble	Interfaz	Reconocimiento de la interfaz	Inflexiones específicas de la interfaz
	Visualización	Síntesis gráfica	Reconocimiento de gestos e intenciones a partir de visualizaciones
	Emoción	Análisis y sincronía y retardos	Disfrute de la interpretación y el ensamble
	Interpretación	Partitura	Apropiación y lectura de la interfaz
Interacción	Interfaz	Objetivos de la interfaz	Modos de control individual y grupal
	Visualización	Visualización de los sensores	La forma, en las visuales, estaba directamente relacionada con la posición del brazo. Percepción visual del gesto
	Emoción	Frustración en control de la interfaz	Conciencia del gesto y su representación en la interfaz
	Interpretación	Lectura de partitura, control del tiempo y sincronía	Interfaz directamente proporcional a las acciones de los intérpretes
Colaboración/colectivo	Interfaz	Lectura de la interfaz	Interpretación
	Visualización	Atención en los detalles	Construcción conjunta, trabajo co-lectivo
	Emoción	Baja colaboración	Actitud participativa y propositiva
	Interpretación	Acuerdos previos	Señales interpretativas y expresivas

Con una serie de técnicas impulsadas por el usuario, nuestro entorno de respuesta evolucionó a través de varias etapas para llegar a su representación actual. Siguiendo las pautas del diseño centrado en el usuario, se comenzó con un enfoque en el usuario mediante observaciones, entrevistas y perfiles de persona que, a su vez, se utilizaron para obtener comentarios de los usuarios mediante testeos formales, en los cuales, el nivel de participación de los usuarios fue en aumento; primero, con la evaluación inicial del prototipo y luego con la colaboración de los laboratorios de análisis, incorporando sus comentarios y sugerencias en cada etapa. Cuestionarios recogidos a lo largo de la colaboración demostraron que nuestro sistema ayudó a mejorar el sentido de los intérpretes en cuanto al disfrute y autoexpresión. Además, el análisis cualitativo de nuestras discusiones con los miembros de los laboratorios, ha demostrado que encontraron el sistema práctico, y que probablemente lo usarían de nuevo.

## **Conclusiones**

En términos generales, se logró el desarrollo de un entorno performático audiovisual telemático, un tele-espacio activo, donde los intérpretes pueden interactuar telemáticamente participando activamente en la construcción del ensamble y experimentar un nuevo campo performático aumentado.

Dentro del desarrollo de los diferentes performances telemáticos se fue categorizando el entorno virtual generado por la confluencia no solo de los espacios remotos sino de las interacciones generadas a partir del que hacer de los intérpretes, los creadores y los técnicos, de esta forma el concepto de espacio en el espacio activo refiere un lugar virtual que es designado o define un nuevo tipo de espacio simbolizado, dicho espacio se conforma teniendo en cuenta lo que se percibe y se desarrolla en el gracias a la relación comunicativa generada por la interacción entre los participantes, es así como esta información perceptual del espacio se construye a partir de la conformación de la interfaz y de la interacción sobre la misma, en el caso del performance se da gracias a la gestualidad del interprete virtualizado a partir del uso de sensores y de su representación en la interfaz embebida (o espacio mediatizado), generándose así un concepto de presencialidad que se desarrolla a partir de la interacción gestual y simbólica sobre la interfaz y en últimas sobre la construcción misma del espacio activo. Esta mirada de la conformación del espacio telemático y de la potencia del gesto como presencia en la construcción de la interfaz es una de las conclusiones laterales más importantes planteadas en esta investigación.

Entender el lugar como el espacio del ensamble, donde los participantes tejen una estructura comunicativa y sobretodo social y donde se establecen tanto lazos como vínculos interpretativos, performáticos y afectivos, brinda al tele-espacio activo una característica de entorno más allá de lo mediatizado y mucho más complejo e interesante que el simple hecho de la telecomunicación, el compartir esta construcción de lugar es uno de los puntos de vista que aporta esta investigación frente a la construcción del performance telemático musical.

El uso de una pieza de creación musical como un experimento que valida las hipótesis de la investigación dentro de un proceso estructurado como un entorno empírico analítico, es una propuesta que, si bien hace parte del concepto investigación creación, se aleja de la estructura comúnmente dispuesta en la que el producto de investigación es una obra de creación, al hacer uso de la creación de obras dentro del proceso de investigación y no únicamente como un producto final se plantea, a partir de esta investigación, una interesante alternativa metódica en el campo de la investigación creación, que puede ser utilizada en el desarrollo de diferentes investigaciones del Doctorado.

## **Agradecimientos**

Este trabajo ha sido financiado parcialmente por el Proyecto de Investigación "Espacios Telemáticos" y forma parte de la Tesis doctoral "Configuración de tele-espacios activos. Entornos de interacción performáticos distribuidos de creación sonora y visual".

## **Referencias**

1. Bernsen, N., & Dybkjær, L. (2009). *Multimodal Usability*. Verlag, London, UK: Springer.
2. Doering, T., Pfleging, B., Kray, C., & Schmidt, A. (2010). Design by Physical Composition for Complex Tangible User Interfaces. *Proceeding CHI EA '10 CHI '10 Extended Abstracts on Human Factors in Computing Systems* (pp. 3541-3546). New York: ACM Publisher.
3. Jordà, S. (2002). FMOL: Toward User-Friendly, Sophisticated New Musical Instruments. *Computer Music Journal*, 26(3), 23-39.
4. MacDonald, C., & Atwood, M. (2013). Changing perspectives on evaluation in HCI: past, present, and future. *CHI EA '13 CHI '13 Extended Abstracts on Human Factors in Computing Systems* (pp. 1969-1978). New York: Publisher ACM.
5. Padhye, N., & Kalia, S. (2009). Rapid prototyping using evolutionary approaches. *Proceeding GECCO '09 Proceedings of the 11th Annual Conference Companion on Genetic and Evolutionary Computation Conference: Late Breaking Papers* (pp. 2725-2728). New York: ACM Publisher.
6. Wood, L. (1997, March). Semi-structured interviewing for user-centered design. *Magazine interactions*, 4(2), 48-61.



## **Capítulo 2**

### **Analítica de datos e inteligencia artificial en la Ciencia**



## Implementación de un sistema experto basado en inteligencia computacional híbrida para ayuda diagnóstica en la especialidad de psiquiatría en pacientes que sufren depresión

Cristian Camilo Carmona Gallego<sup>a</sup>, Gustavo Adolfo Isaza Echeverry<sup>b</sup>, Oscar Mauricio Castaño  
Ramírez<sup>c</sup>

<sup>a</sup>Grupo de Telesalud - Facultad de Ciencias Para la Salud - Universidad de Caldas,  
Cra. 25 No 48 – 57 (Manizales, Colombia).  
cristian.camona\_g@ucaldas.edu.co

<sup>b</sup>GITIR, Facultad de Ingenierías – Universidad de Caldas,  
Cl. 65 #26-10 (Manizales, Colombia).  
gustavo.isaza@ucaldas.edu.co

<sup>c</sup>Grupo Salud mental y comportamiento Humano - Facultad de Ciencias Para la Salud - Universidad de  
Caldas,  
Cra. 25 No 48 – 57 (Manizales, Colombia).  
oscar.castano@ucaldas.edu.co

**Resumen.** El grupo de investigación Telesalud de la Universidad de Caldas ha mejorado constantemente la prestación de servicios de telemedicina por medio de sus diferentes plataformas permitiendo prestación de los servicios de Telemedicina en 17 especialidades médicas y realizando más de 35.000 actividades de telemedicina. Uno de los últimos avances se generó a partir del desarrollo de una investigación de la maestría en ingeniería computacional de la Universidad de Caldas, con el apoyo de los grupos de investigación Telesalud, Grupo de Investigación en Tecnologías de la Información y Redes (GITIR), y Grupo de investigación de Salud mental y comportamiento Humano. El desarrollo consiste en la integración de un módulo basado en inteligencia computacional híbrida y la plataforma de servicios de telemedicina para la ayuda diagnóstica en pacientes que sufren depresión. La implementación de este proyecto se tuvieron en consideración diferentes algoritmos como son bosques aleatorios, redes neuronales, bayesiano, regresión logística y máquinas de vectores de soporte (SVM). Además se cuenta con la representación, vinculación y distribución basada en ontologías con LOD (Linked Open Data) para mostrar de manera semántica los factores más determinantes en el diagnóstico de una depresión. Uno de los principales objetivos de este proyecto es la mejora de la prestación de servicios especializados de salud teniendo en cuenta que en la actualidad el número de psiquiatras con los que cuenta Colombia es muy bajo.

**Abstract.** Telesalud research group of the University of Caldas has constantly improved the provision of telemedicine services through its different platforms allowing the provision of telemedicine services in 17 medical specialties and carrying out more than 35,000 telemedicine activities. One of the latest advances was generated from the development of an investigation of the master's degree in computer engineering from the University of Caldas, with the support of the Telesalud research

group, Grupo de Investigación en Tecnologías de la Información y Redes (GITIR), and Grupo de investigación de Salud mental y comportamiento Humano. The development consists in the integration of a module based on hybrid computational intelligence to the telemedicine services platform for diagnostic support in patients suffering from depression. Different algorithms such as random forests, neural networks, Bayesian, logistic regression and support vector machines (SVM) were considered in the development of this project. Furthermore, there is representation, linking and distribution based on ontologies with LOD (Linked Open Data) to semantically show the most determining factors in the diagnosis of depression. One of the main objectives of this project is the improvement of the provision of specialized health services, taking into account that currently the number of psychiatrists that Colombia has is very low.

**Palabras Clave:** Linked Open Data – inteligencia artificial - analítica de datos – sistemas inteligentes híbridos – Web Semántica – psiquiatría – Depresión – Ayuda diagnóstica.

**Eje temático:** e-salud Gestión de datos y datos enlazados (Linked Data).

## **Introducción**

Debido a que Colombia presenta un déficit en el número de especialistas en psiquiatría, el cual es de 1 por cada 100.000 habitantes [1], surge la necesidad de crear sistemas inteligentes que ayuden a suplir esta necesidad, para mejorar la atención del servicio especializado de salud y a su vez mejorar la calidad de vida de los pacientes atendidos.

La Especialidad de psiquiatría cuenta con un manual diagnóstico y estadístico de los trastornos mentales, denominado el DSM5 [2] en cual se tienen categorizados los diferentes tipos de trastornos psiquiátricos, y como se pueden clasificar a partir de distintos criterios y síntomas que presenta un paciente.

En recientes décadas ha surgido el interés de la investigación en sistemas inteligentes asociados al campo de la medicina, mostrando una serie de exitosos proyectos y aplicaciones para el procesamiento de imágenes, generación de diagnósticos en enfermedades específicas creación de sistemas de apoyo de decisión y predicción de resultados.[3].

Teniendo en cuenta lo anterior, este proyecto se basó diseñar y desarrollar un prototipo sistema experto basado en inteligencia artificial híbrida para la generación de ayudas diagnósticas que apoye a la toma de dediciones de los médicos cuando traten a pacientes que sufren depresión.

Para el desarrollo e implementación del proyecto se contó con el apoyo de los grupos de investigación Telesalud, grupo GITIR y el grupo Salud mental y comportamiento Humano, todos pertenecientes a la Universidad de Caldas.

## **Problemática**

La prestación de los servicios de salud de calidad es una de las necesidades primordiales de la sociedad colombiana, y dentro de estos servicios está presente la psiquiatría, la cual es la rama de la ciencia médica que se ocupa del cerebro y el comportamiento humano [4], sin embargo, hoy en día existe una crisis en el número de especialistas en el campo de la psiquiatría, en el cual según un estudio realizado por la Pontificia Universidad Javeriana se muestra los siguientes datos, en Colombia existen 687 Psiquiatras de los cuales hay un psiquiatra disponible por cada 100.000 habitantes. Otros datos interesantes que muestra el estudio es la dificultad de vinculación de psiquiatras la cual es de un 70% mientras que la capacidad resolutive de este tipo de consultas es del 68% [1].

Lo anterior conlleva que muchos de los pacientes que deben ser tratados en psiquiatría no obtienen un diagnóstico oportuno y apropiado para su enfermedad, lo cual puede significar una complicación para la enfermedad que padece el paciente y a su vez desmejorar la calidad de vida de su familia. Por otro lado hay que mencionar que muchos de los pacientes viven en zonas alejadas y si desean ser tratados por un psiquiatra deben realizar viajes muy largos, lo cual genera costos muy altos para su tratamiento.

Una alternativa para que el gobierno nacional aumente su presencia en las zonas alejadas del Colombia que tienen servicios especiales de salud deficientes, es la telemedicina, la cual es definida por la resolución 2003 de 2014 del Ministerio de Salud

y Protección Social de Colombia, como “la modalidad de prestación de servicios de salud, realizados a distancia, en los componentes de promoción, prevención, diagnóstico, tratamiento y rehabilitación, por profesionales de la salud que utilizan tecnologías de la información y la comunicación, que les permiten intercambiar datos con el propósito de facilitar el acceso y la oportunidad en la prestación de servicios de salud a la población que presenta limitaciones de oferta, de acceso a los servicios o de ambos en su área geográfica” [5]. No obstante, la telemedicina se ha demostrado que es efectiva, pero es muy costosa, por lo tanto, las entidades prestadoras de servicios de salud no han podido masificar esta práctica que puede mejorar sustancialmente el acceso a servicios médicos especializados.

Si bien el acceso a la tecnología es cada vez más fácil y muchas instituciones prestadoras de servicios están dotadas tecnológicamente, en este momento no se encuentran referencias que evidencien la implementación de sistemas expertos que ayuden a la toma de decisiones al momento de realizar una consulta médica en psiquiatría, algunas posibles causas pueden ser:

- Poca capacitación de los médicos en el uso de este tipo de sistemas.
- La deficiencia en la interoperabilidad entre sistemas médicos ya que en muchas ocasiones las IPS cuentan con muchos sistemas de información que cumplen la misma función, pero no están conectados entre sí, por lo tanto, existe redundancia de la información.
- La desconfianza que puede generar el uso de sistemas expertos para la recomendación diagnóstica.
- Malas redes de comunicaciones.

Si bien los médicos generales son quienes atienden en primer instancia a los pacientes que tienen una enfermedad psiquiátrica, la percepción de los pacientes frente a su preparación los médicos para estos temas es del (32.5%) [según 1 tablas 17 y 19], es decir hace falta también más capacitación para los médicos generales.

Hay que mencionar que la producción bibliográfica en Colombia para temas de telemedicina es muy poca, según el portal Publindex desde el año 2001 hasta el año 2016 hay 37 artículos, y para el tema de telepsiquiatría existe solo 1 artículo.

El grupo de investigación Telesalud de la Universidad de Caldas ha sido pionero en temas de telemedicina y temas de telepsiquiatría, como es el proyecto de investigación telepsiquiatría en cárceles, además cuenta con una plataforma de servicios de telemedicina que se encuentra en su cuarta versión con más de 15 años de desarrollo y que ha permitido tratar a cerca de 13.000 pacientes en distintas especialidades. Sin embargo hasta ahora solo se ha realizado investigación en el modelo de telemedicina tradicional y no se ha explorado la implementación de inteligencia artificial.

Los sistemas de razonamiento basados en casos (CBR por sus siglas en inglés), combinados con otro método de inteligencia artificial que permita la inferencia de conocimiento, son muy usados en sistemas de salud mostrando una muy buena eficiencia. Entre 77 sistemas CBR revisados, ninguno se combina con ontologías que permitan modelar el conocimiento y solo existe un sistema que aborda el tema de la depresión [6].

Si bien las ontologías están presentes en diferentes proyectos, aún estas no muestran gran presencia en el campo de sistemas inteligentes con fines médicos, y es por esto que se ha decidido buscar la combinación de diferentes modelos basados en inteligencia artificial que combinados con ontologías, ayuden generar ayudas diagnósticos en el

campo de la psiquiatría para pacientes que sufren depresión.

## **Justificación**

Con el sistema desarrollado se muestra la eficiencia del uso de datos enlazados y ontologías combinadas con sistemas inteligentes híbridos, las cuales no han sido muy exploradas, además la misma ontología sería un sub resultado de la investigación. Este proyecto aporta nuevo conocimiento a la manera en que se puede realizar telemedicina, debido a que el modelo del sistema propuesto necesita que el aprendizaje del sistema sea supervisado por un experto y a su vez permite una nueva versión sobre la plataforma de telemedicina que se utiliza actualmente en el grupo de investigación Telesalud Universidad de Caldas.

Considerando que hay muy poca producción bibliográfica para los temas de telemedicina y telepsiquiatría, este proyecto pretende aumentar los documentos científicos sobre los campos mencionados.

Desde el aspecto tecnológico se espera que tanto médicos generales y especialistas se apropien más de este tipo de tecnología, y además las recomendaciones diagnósticas que genera el sistema puedan ser utilizadas de manera académica por parte los usuarios. Por otro lado el proyecto pretende mostrar a los usuarios, los beneficios del uso de inteligencia artificial para mejorar el servicio que se presta a los pacientes.

Un aporte muy importante con el sistema propuesto es que puede ayudar a suplir el déficit actual que se tiene de especialistas, teniendo en cuenta que a medida que el sistema sea entrenado, cada una de las recomendaciones diagnósticas generada estará basada en los diagnósticos que se han validado en consultas anteriores, aumentando así la probabilidad de éxito de la recomendación generada. Este proyecto puede ayudar a la masificación de servicios de telemedicina. Además, en un largo plazo se puede extrapolar esta metodología a diferentes especialidades que tienen el mismo problema. Cabe mencionar que este tipo de sistemas no sustituye la interacción que se tiene entre médicos y pacientes ya que la responsabilidad es del médico tratante.

Para la implementación de este proyecto se escogió los trastornos de depresión, la cual es una enfermedad de salud pública que es muy común hoy en día, que muchas personas pasan por desapercibido, y al momento que son diagnosticadas pueden estar en un estado avanzado.

## **Descripción de la solución tecnológica implementada**

A continuación se muestra cual es el proceso que tenía antes de implementar el nuevo módulo de inteligencia artificial:

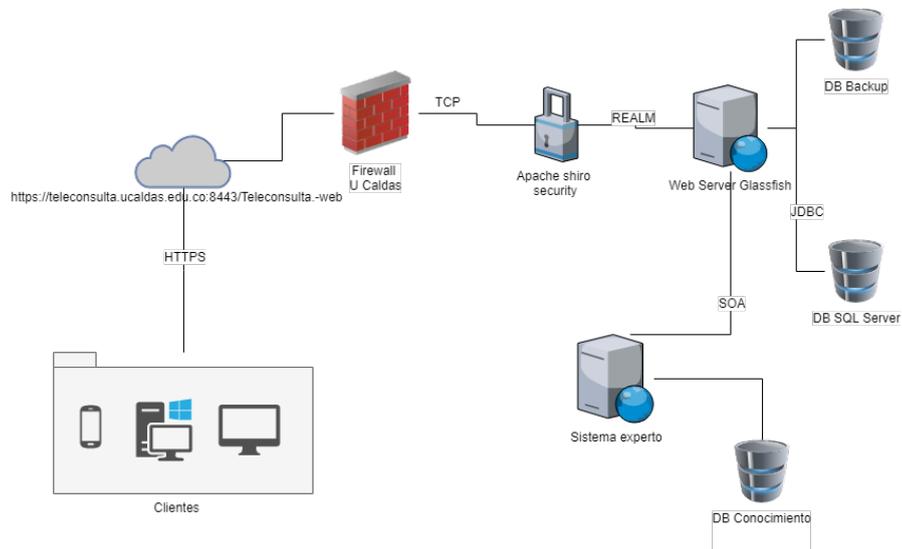
El médico general es quien se encarga de atender un paciente y registrar la información personal y la información correspondiente a una consulta médica. Luego de terminar la consulta médica, el sistema remite la consulta hacia un especialista el cual es notificado que tiene una nueva consulta por responder. El especialista revisa el caso y emite su concepto sobre la consulta realizada. El especialista determina si resuelve el caso por completo, solicita realizar una nueva consulta, o solicita una

remisión para realizar una consulta de manera presencial.

Para la implementación e integración del módulo de inteligencia artificial híbrida con la plataforma de telemedicina se contó con el diseño de un instrumento que permitiera la captura de todos los datos que ayudarían a la clasificación de un paciente con depresión, el instrumento fue diseñado con el apoyo de Oscar Mauricio Castaño Ramírez, Médico Psiquiatra de la Universidad de Caldas.

Cada una de las preguntas del instrumento diseñado están basadas en el manual DSM5, permitiendo acercarse así de una manera más correcta a las diferentes clasificaciones que genera el sistema.

En este momento después de diligenciarse una consulta médica con ayuda del instrumento diseñado, el sistema evalúa las variables y emite una ayuda diagnóstica con la cual el médico decide si es correcta su clasificación, de ser correcta el médico agrega el diagnóstico generado a la consulta, si es incorrecto este nuevo caso ayuda al entrenamiento del algoritmo inteligente.



**Fig. 1.** Representación general de la arquitectura del sistema

La figura anterior nos da una vista de la representación general de la arquitectura implementada en el desarrollo del proyecto

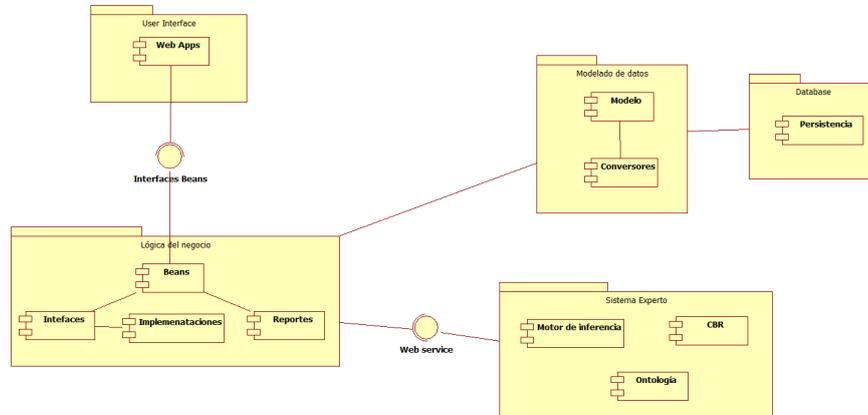


Fig. 2. Vista de desarrollo

**User Interface:** Este paquete se encarga de contener todas las interfaces de usuario del sistema.

**Lógica del negocio:** En este paquete está implementado la mayoría de las funcionalidades del sistema de servicios de telemedicina, las interfaces de usuario se comunican con la lógica del negocio por medio de los beans, también es en los beans es donde se define el tiempo de vida de cada bean. Los beans invocan las funcionalidades de las implementaciones por medio de las interfaces las cuales contienen el llamado de los métodos, y también llaman algunas clases que generan los reportes. Las implementaciones en su gran mayoría son interacciones con la base de datos por medio de ORM.

**Modelado de datos:** Aquí se encuentra el modelo de los datos, los cuales han sido mapeados desde la base de datos, y algunos conversores que sirven para el manejo de objetos desde las interfaces.

**Database:** Se encuentra la persistencia del sistema transaccional.

**Sistema experto:** Este módulo contiene el desarrollo del sistema experto el cual se comunicará con el sistema de servicios de telemedicina por medio de un web service.

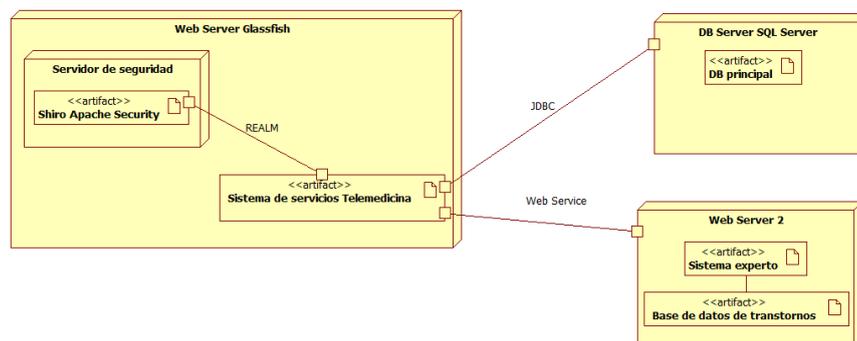


Fig. 3. Vista Física

El web server glassfish, es el servidor donde se encuentra desplegado el sistema de servicios de telemedicina, dentro del existe un servidor de seguridad llamado Apache shiro security que es el que se encarga de gestionar el acceso de usuarios. El servidor de seguridad se comunica con la plataforma de servicios de telemedicina mediante un REALM el cual se comunica con la base de datos y obtiene los tokens de ingreso de los usuarios.

El siguiente nodo es donde se encuentra el sistema de base de datos, allí está desplegada la base de datos principal, la cual está instalada en un motor de base de datos SQL Server, la plataforma de telemedicina se comunica con el motor de base de datos, mediante un controlador JDBC.

El último nodo es el encargado de almacenar el sistema experto, el cual comparte la funcionalidad de generar un diagnóstico a por medio de un web service.

## Resultados

- Los algoritmos que hasta el momento mostrado mejor desempeño son máquinas de vectores de soporte por sus siglas en inglés (SVM) y algoritmos bayesianos, los cuales muestran un porcentaje de acierto del 92 % y 90 % respectivamente con desviación de resultados de 5%.
- El sistema implementado se considera que aún se encuentra en etapa de experimentación o prototipado y está sujeto a cambios a medida que se obtengan más datos, en el momento se cuenta con una clasificación exitosa de 40 pacientes.
- Los algoritmos han sido validados bajo las mismas características utilizando el método de validación cruzada.
- Ya se tiene una ontología base, la cual está en proceso de constante mejora y validación.

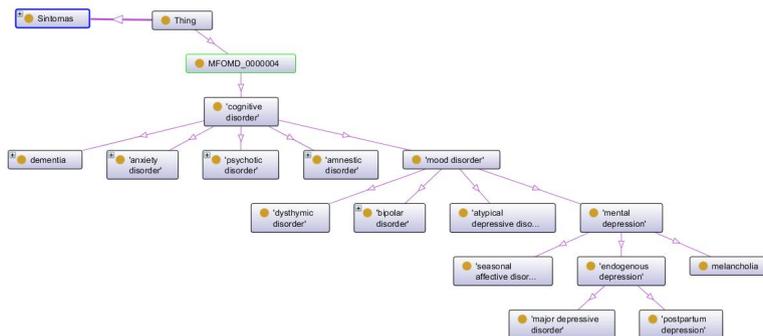


Fig 4. Representación gráfica de la ontología, sección enfermedades mentales

La anterior figura contiene la ontología donde se representan la clasificación de enfermedades mentales, para motivos de este proyecto se clasifica (mayor depressive disorder).

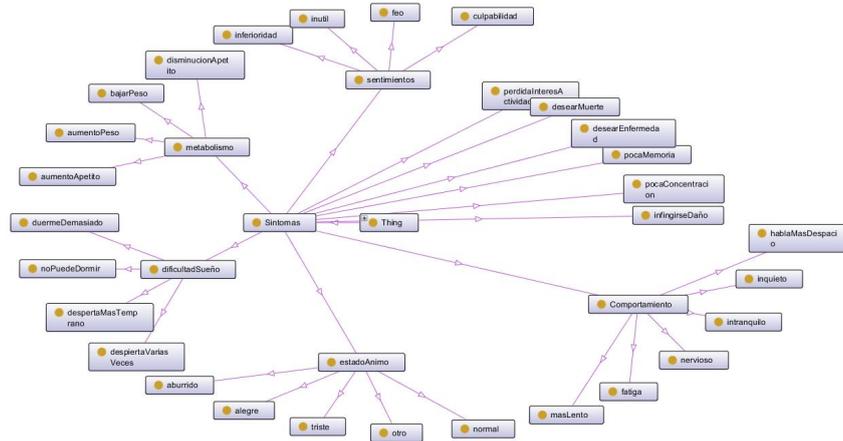


Fig. 5. Representación gráfica de la ontología, sección instrumento.

En la gráfica anterior se tiene el grafo donde se muestran las reglas que están asociadas al instrumento diseñado para la captura de datos.

## Impactos

A continuación se muestran distintos impactos esperados

**Impactos tecnológicos:** Se ha logrado la integración exitosa del módulo inteligente con el sistema transaccional de servicios de telemedicina.

Se ha creado una ontología base que está basada en estándares de Linked Open Data.

A mediano plazo el impacto es repetir el modelo propuesto de sistemas expertos para la generación de ayudas diagnósticas en otros trastornos y especialidades.

A largo plazo adoptar en las políticas públicas de telemedicina el uso de sistemas expertos como una alternativa a los modos de realizar telemedicina existente.

**Impactos Económicos:** El impacto generado con este proyecto es la de generar un factor diferenciador en la prestación de servicios de telemedicina por parte del grupo Telesalud Universidad de Caldas, mejorando así el proceso con el que se realiza actualmente la telemedicina.

**Impactos Sociales:** Con este proyecto se busca subsanar el déficit que existe actualmente de especialistas en psiquiatría en el país, además de continuar contribuyendo en mejorar la calidad de vida de los pacientes que son atendidos por una consulta especializada en telemedicina.

## **Aprendizajes**

- Se evidencia que este tipo de soluciones pueden ayudar a mejorar la consulta primaria ya que como se mencionó, Colombia presenta una crisis de especialistas, y mediante la telemedicina esto puede solucionar una parte de la problemática en sitios que presenta un principio de oportunidad.
- Debido a que el sistema de salud colombiano muestra un aumento en el uso de historia clínica electrónica, esto representa una oportunidad para establecer sistemas que permitan la gestión de conocimiento desde diferentes fuentes, con la finalidad de toma de decisiones más costo efectivas
- Se ha encontrado en las revisiones bibliográficas que la implementación de ontologías combinadas con sistemas expertos en el campo de la salud no ha sido explorada y por tal motivo se encuentra un vacío de conocimiento que puede ser explotado por esta clase de proyectos.
- El modelo propuesto y su arquitectura puede ser replicado en más campos de la medicina que tienen iguales dificultades.

## **Agradecimientos**

La facultad de ciencias de la salud de la Universidad de Caldas y la gestión de su Decana Dra. Claudia Patricia Jaramillo.

El grupo Telesalud y su grupo de profesionales que de la mano de su coordinador Esteban Granada han permitido el desarrollo de este tipo de iniciativas que ayudan a salvar vidas.

Agradecemos a Red RENATA Red Nacional Académica de Tecnología Avanzada por la infraestructura de red de tecnología avanzada y conexión de alta velocidad que permite desde las universidades el desarrollo de proyectos colaborativos de ciencia, educación e innovación al igual que una conexión estable y segura para los participantes virtuales en el continente latinoamericano, y a la RedClara por realizar convocatorias que incentivan y hacen posible que las personas que integran las universidades construyan e innoven en beneficio de la comunidad académica iberoamericana.

## **Referencias**

1. Amaya, J., et al., Estudio de disponibilidad y distribución de la oferta de médicos especialistas, en servicios de alta y mediana complejidad en Colombia Documento Técnico GPES/1682C-13. 2013, Universidad Pontificia Bolivariana.
2. Association, A.P., Manual diagnóstico y estadístico de los trastornos mentales DSM-5. 5 ed. 2014: Editorial Médica Panamericana.
3. Shaikhina, T. and N.A. Khovanova, Handling limited datasets with neural networks in medical applications: A small-data approach. *Artificial Intelligence in Medicine*, 2017. **75**: p. 51-63.

4. Singh, P., A.P. Singh, and S. Ahmad. Case based reasoning model in the diagnosis of psychiatric disorder. in 2016 International Conference on Communication and Electronics Systems (ICCES). 2016.
5. minsalud. Resolución 2003 de 2014. 2014 [cited 2017 Octubre 2017]; Available from:  
[https://www.minsalud.gov.co/Normatividad\\_Nuevo/Resoluci%C3%B3n%202003%20de%202014.pdf](https://www.minsalud.gov.co/Normatividad_Nuevo/Resoluci%C3%B3n%202003%20de%202014.pdf).
6. Choudhury, N. and S. Ara, A Survey on Case-based Reasoning in Medicine. Vol. 7. 2016.

## Uso de metaheurísticas en la segmentación de mamografías para el prediagnóstico de cáncer de mama

David González Patiño<sup>a</sup>, Yenny Villuendas Rey<sup>b</sup>, Amadeo J. Argüelles Cruz <sup>a</sup>,

<sup>a</sup> Centro de Investigación en Computación del Instituto Politécnico Nacional, Avenida Juan de Dios Bátiz esq. Miguel Othón de Mendizábal, Nueva Industrial Vallejo, Gustavo A. Madero, CP 07738, Ciudad de México, D.F., México  
davidglezp-92@hotmail.com, aarguelles@ipn.mx

<sup>b</sup> Centro de Innovación y Desarrollo Tecnológico en Cómputo del Instituto Politécnico Nacional, Av. Juan de Dios Bátiz s/n, Nueva Industrial Vallejo, Gustavo A. Madero, 07700 Ciudad de México, D.F., México  
yenny.villuendas@gmail.com

**Resumen.** En este trabajo se realiza la experimentación del uso de metaheurísticas aplicadas en la segmentación de mamografías. La mayoría de las metaheurísticas son diseñadas para resolver problemas de optimización, sin embargo en este trabajo se presenta su utilización para realizar la segmentación utilizando el Índice de Dunn como función de optimización. Los resultados mostraron un desempeño prometedor de las metaheurísticas comparadas con respecto a dos algoritmos clásicos de segmentación.

**Palabras Clave:** segmentación, mamografías, cáncer de mama, metaheurísticas, optimización

**Eje temático:** e-Salud

## **1. Introducción**

El cáncer de mama es un problema de salud actual en todo el mundo, es por esto que surge la necesidad de realizar un diagnóstico temprano del mismo, logrando así la supervivencia del paciente. Este tipo de cáncer ha causado la muerte de más de 627,000 contadas hasta 2018 y se estima un aumento en todo el mundo y sobre todo en países subdesarrollados [1]. De igual forma se estima la muerte de 42,260 personas (41,760 mujeres y 500 hombres) únicamente en 2019.

El uso de mamografías permite obtener una imagen de la mama en escala de grises, la cual se puede analizar para identificar lesiones o posibles microcalcificaciones. Es por este motivo que es necesario contar con algoritmos que permitan analizar la imagen mamográfica mediante la segmentación, que consiste en dividir la imagen en grupos de píxeles con el objetivo de ser más fácil de implementar [2] y en consecuencia servir de apoyo en el diagnóstico de la enfermedad.

Una buena segmentación debe tener características uniformes y homogéneas de tal forma que los límites sean simples y las regiones adyacentes posean características diferentes [3]. Actualmente no existe un algoritmo perfecto que realice la tarea de segmentación de forma correcta para cada caso, esta es la motivación por la que en este trabajo se propone utilizar metaheurísticas para la segmentación de imágenes de mamografías.

Las metaheurísticas son algoritmos que imitan los comportamientos de la fauna, así como procesos biológicos [4]. Gran parte de estos algoritmos están diseñados y son utilizados para resolver problemas de optimización debido a su bajo coste computacional y su comportamiento estocástico. En este artículo se realiza una experimentación sobre el uso de metaheurísticas aplicadas en la segmentación de imágenes mamográficas.

Este artículo está organizado de la siguiente forma: En la sección 2 se presenta la problemática y el contexto en que se propuso esta investigación realizada. En esa misma sección se presentan algunos de los trabajos previamente realizados. La descripción de la solución se expone en la sección 3. En la sección 4 se presentan los aspectos relevantes en de la experimentación realizada. Los resultados obtenidos son expuestos en la sección 5. Y finalmente en la sección 6 se presentan las conclusiones y trabajo a futuro.

## **2. Problemática y contexto**

La segmentación es un proceso que consiste en dividir las Regiones de Interés (Regions of Interest, ROI) y el fondo en una imagen. Para realizar este proceso existen muchas técnicas, las cuales de acuerdo con Fu et al. [5] pueden ser categorizadas en tres tipos: detección de bordes, extracción de regiones y agrupamiento.

La detección de bordes es una técnica basada en identificar los bordes de una región con el objetivo de tener un drástico cambio en la escala de grises de cada región adyacente. Esta técnica ha sido utilizada por Dollár y Zitnick [6], así como Malik et al. [7] en años recientes. De igual forma se han propuesto algoritmos para mejorar el contraste y lograr definir los bordes de forma más acertada, como en el trabajo de

Mencattini et al. [8]. Esta técnica es muy útil para detectar discontinuidades, así como propiedades en los materiales e iluminación.

La extracción de regiones es un método que aporta más información que la técnica anterior, puesto que una región tiene más píxeles que proporcionan información sobre la región. Generalmente se utiliza una combinación de ambos algoritmos (detección de bordes y extracción de regiones) para obtener la mayor cantidad de información sobre una región de interés.

Finalmente el agrupamiento, también llamado *clustering*, es una técnica basado en determinar los píxeles en distintos clústeres. Algo que cabe resaltar es que no necesariamente los píxeles deben estar juntos en la imagen. Este enfoque ha sido utilizado para la segmentación de peces [9] y células [10], entre otros.

Una de las notables ventajas de este enfoque es que los algoritmos son fáciles de implementar y simples, por lo que los resultados son fáciles de interpretar. Los algoritmos de agrupamiento son sensibles a los valores anormales y el resultado depende en gran medida de los puntos iniciales. Un aspecto que puede ser visto como ventaja o desventaja es la capacidad del algoritmo de *clustering* para asignar todos los elementos a un grupo.

Los algoritmos utilizados en este trabajo corresponden a la tarea de agrupamiento.

Uno de los algoritmos clásicos y actualmente utilizado de segmentación es el método de umbralado de Otsu [11]. Este método es muy utilizado por su simplicidad y buen desempeño. Dicho algoritmo se basa en encontrar el umbral óptimo maximizando o minimizando las varianzas entre los elementos fuera y dentro de cada clúster respectivamente. Este método ha sido ampliamente utilizado en trabajos recientes [12–14].

Existen muchas técnicas de segmentación, sin embargo no existe una teoría o método para determinar cuál es el mejor algoritmo para segmentación de imágenes digitales [15]. Esta es la motivación para realizar una investigación del uso de metaheurísticas en la segmentación de imágenes mamográficas.

### 3. Descripción de la solución

En este trabajo se proponen cuatro algoritmos basados en comportamientos biológicos, de los cuales se presenta su información y detalles en esta sección. Estos algoritmos fueron comparados con el algoritmo Otsu, el cual fue presentado en la sección anterior.

Los algoritmos evaluados en este trabajo son: Algoritmos genéticos (*Genetic Algorithms*, GA), Comportamiento de los murciélagos y ecolocalización (*Novel Bat Algorithm*, NBA), Optimización de malezas invasivas (*Invasive Weed Optimization*, IWO) y Optimización de enjambre de partículas (*Particle Swarm Optimization*, PSO).

Los algoritmos genéticos (GA) son metaheurísticas utilizadas para optimización que se inspiran en la evolución natural. Estos algoritmos se basan en la mutación, cruzamiento y selección de los individuos [16]. Estos algoritmos utilizan una población para encontrar posibles soluciones buscando encontrar la mejor solución en un problema de optimización. Esta población es generada aleatoriamente al comienzo y utiliza mutación, cruzamiento y operadores de selección para encontrar la mejor solución en cada iteración.

*Bat algorithm* (BA) es un algoritmo propuesto en 2012 [17], el cual resuelve problemas de optimización. Este algoritmo fue modificado en 2015 [18] (NBA) incluyendo nuevas características como el efecto Doppler. La fundamentación de este algoritmo se basa en el comportamiento de los murciélagos y la ecolocalización que utilizan para localizar e identificar objetos en un espacio determinado. Estos murciélagos virtuales utilizan la ecolocalización para medir la distancia y modificar su velocidad y frecuencia con el objetivo de buscar posibles soluciones para la función de desempeño que se pretende optimizar.

Otra metaheurística utilizada en este trabajo fue el algoritmo basado en la colonización de hierbas conocido como *Invasive Weed Optimization* (IWO). Este algoritmo numérico estocástico es un algoritmo de optimización el cual usa la adaptación del ambiente y la proliferación de las hierbas. Fue ideado por Mehrabian y Lucas en 2006 [19] y se inspira en la propagación de hierbas invasivas en la naturaleza. Un aspecto importante de este algoritmo es que todas las plantas participan en las nuevas generaciones produciendo semillas que producirán más plantas. La cantidad de semillas que una planta produce estará limitada por el desempeño de dicha planta con respecto a la función de optimización. Cada planta se reproduce independientemente sin necesidad de relacionarse con otras.

Finalmente pero no menos importante, modificamos el algoritmo *Particle Swarm Optimization* (PSO) para realizar la segmentación de las imágenes digitales. Este algoritmo fue propuesto en 1995 por Kennedy y Eberhart [20] basándose en el comportamiento colaborativo de aves y peces mientras buscan comida. Este algoritmo es bastante simple y por la misma razón su costo computacional es menor comparado con otras metaheurísticas más complejas. Sin embargo tiene una gran robustez y alta eficiencia computacional.

Es importante comprender que la mayoría de las metaheurísticas son diseñadas y utilizadas en problemas de optimización. Este es el motivo por el cual en este trabajo se plantea la posibilidad de utilizarlas para realizar segmentación de mamografías.

#### **4. Aspectos relevantes de la experimentación realizada**

El banco de datos (Breast Cancer Digital Repository) [21] utilizado en este trabajo fue proporcionado por la Facultad de Medicina de la Universidad de Porto en Portugal. Este banco de datos consta de 200 lesiones de 190 mujeres portuguesas, de igual forma contiene 359 segmentaciones realizadas por expertos radiólogos lo cual constituye las pruebas de verdad (*Ground truth*) que son utilizadas para verificar la correcta segmentación utilizando los algoritmos propuestos. En este trabajo se utiliza el Índice de Dunn como función de optimización, debido a que las metaheurísticas son mayormente utilizadas para optimización. El Índice de Dunn [22] es utilizado para evaluar algoritmos de agrupamiento, mientras mayor sea el Índice de Dunn, mejor será el agrupamiento. Este índice es calculado como la distancia mínima entre elementos de grupos ajenos sobre la distancia máxima entre elementos internos del grupo.

Para esto se define cada uno de los individuos como un conjunto de niveles de gris, el cual nos permitirá encontrar un umbral para separar cada grupo. En la figura 1 se presenta un ejemplo de un individuo con 3 componentes y cada valor del componente

representa un nivel en la escala de grises.

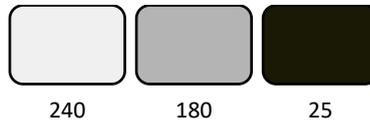


Fig. 1. Imagen de representación de un individuo con 3 componentes

En la figura 2 se muestra un ejemplo visual de 3 segmentaciones producidas por cada individuo, es importante aclarar que cada componente de los individuos representa un nivel de gris por lo que la segmentación se realizará asignando cada pixel al componente (nivel de gris) más cercano.

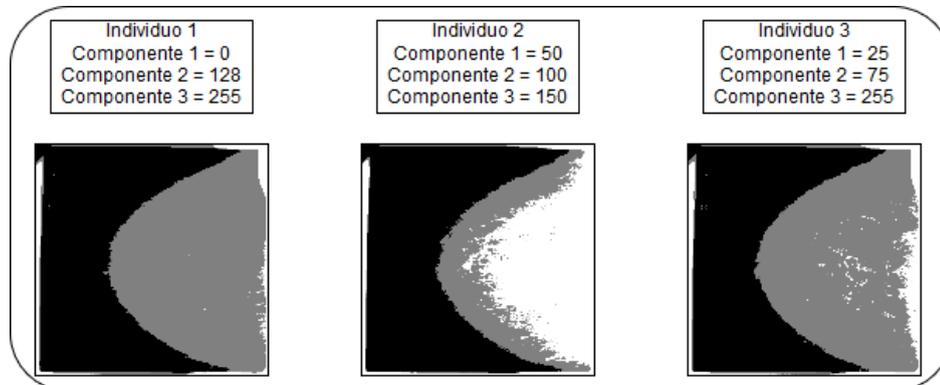


Fig. 2. Segmentaciones producidas por tres individuos independientes con diferentes componentes

Con el uso del Índice de Dunn, presentado en la ecuación 1, es posible utilizar las metaheurísticas propuestas como algoritmos de agrupamiento y por consiguiente utilizarlos en la segmentación de cáncer de mama en mamografías.

$$f = \frac{\text{Mínima distancia entre pixeles de diferente grupo}}{\text{Máxima distancia entre pixeles del mismo grupo}} \quad (1)$$

La función de optimización nos permite calcular la dispersión entre los pixeles de grupos diferentes y pixeles del mismo grupo. Es por este motivo que se busca maximizar esta función, dado que se prefiere que el numerador sea un valor muy grande y el denominador sea un valor muy pequeño. La distancia utilizada es la distancia euclidiana.

Estos algoritmos fueron utilizados con parámetros de prueba, sin embargo la modificación automática de estos parámetros es un factor que puede lograr un mejor desempeño del algoritmo.

## 5. Resultados obtenidos y aprendizajes

En esta sección se presentan los resultados de cada algoritmo aplicado a las 359 imágenes mamográficas.

Para lograr un análisis objetivo de los resultados es necesario utilizar una medida de desempeño del algoritmo. En este trabajo se utilizan dos medidas de desempeño para evaluar distintos aspectos de la segmentación.

La medida de Pureza fue propuesta por Zhao y Karypis en 2001 [23] y es utilizada actualmente debido a su simplicidad [24]. De acuerdo con lo reportado en dicho artículo, esta medida no está sesgada cuando la segmentación tiene pocos grupos. Esta medida es conveniente dado que en este trabajo tenemos únicamente dos grupos: lesión y fondo.

La pureza utiliza la cantidad de objetos (N), los grupos en el clúster de la imagen segmentada (k), los grupos en clúster de la imagen real (*Ground Truth*) y la cantidad de objetos del grupo g que pertenecen al grupo k ( $N_k^g$ ). La ecuación del cálculo de la pureza se presenta en la ecuación 2.

76

$$Pureza = \sum_{k \in Cluster} \frac{1}{N} \max_{g \in GroundTruth} \{N_k^g\}. \quad (2)$$

De forma similar se utiliza la medida F propuesta por Fung en 2003 [25], la cual es ampliamente utilizada [26] y se basa en el cálculo de la precisión y sensibilidad.

Para entender los siguientes cálculos es necesario definir el concepto de Falso positivo (FP), Falso negativo (FN) y Verdadero Positivo (VP), tal como se muestra en la Tabla 1.

**Tabla 1.** Matriz de confusión para Segmentación

		Agrupamiento verdadero	
		Píxeles Negros	Píxeles Blancos
Agrupamiento propuesto por el algoritmo	Píxeles Negros	Verdadero o Positivo	Falso Positivo
	Píxeles Blancos	Falso Negativo	Verdadero o Negativo

Una vez identificados estos indicadores, se procede a definir las ecuaciones para calcular la medida de F.

$$\text{Precisión} = \frac{TP}{TP + FP} \quad (3)$$

$$\text{Sensibilidad} = \frac{TP}{TP + FN} \quad (4)$$

Utilizando las ecuaciones 3 y 4, se realiza el cálculo de F de acuerdo con las ecuaciones 5 y 6.

$$F(m, k) = \frac{2 * \text{Sensibilidad}(m, k) * \text{Precisión}(m, k)}{\text{Sensibilidad}(m, k) + \text{Precisión}(m, k)} \quad (5)$$

$$F = \sum_{m \in GTSG} \frac{|m|}{N} \max_{k \in ASG} F(m, k) \quad (6)$$

Los cálculos fueron realizados para las 359 imágenes comparando la segmentación generada por cada algoritmo con su prueba de verdad (*Ground truth*) de tal forma que es posible obtener un promedio de cada segmentación.

Dichos promedios se presentan en las Tablas 2 y 3.

**Tabla 2.** Pureza media en algoritmos de segmentación

Algoritmo	Pureza media
<b>GA</b>	<b>98.571</b>
IWO	98.480
NBA	98.541
OTSU	98.526
PSO	98.446

Con respecto a la Tabla 2 podemos determinar que dos de las cuatro metaheurísticas propuestas obtuvieron un mejor desempeño que el algoritmo Otsu, lo cual es bastante interesante considerando que el algoritmo OTSU es un método ampliamente utilizado por su simplicidad y buen funcionamiento.

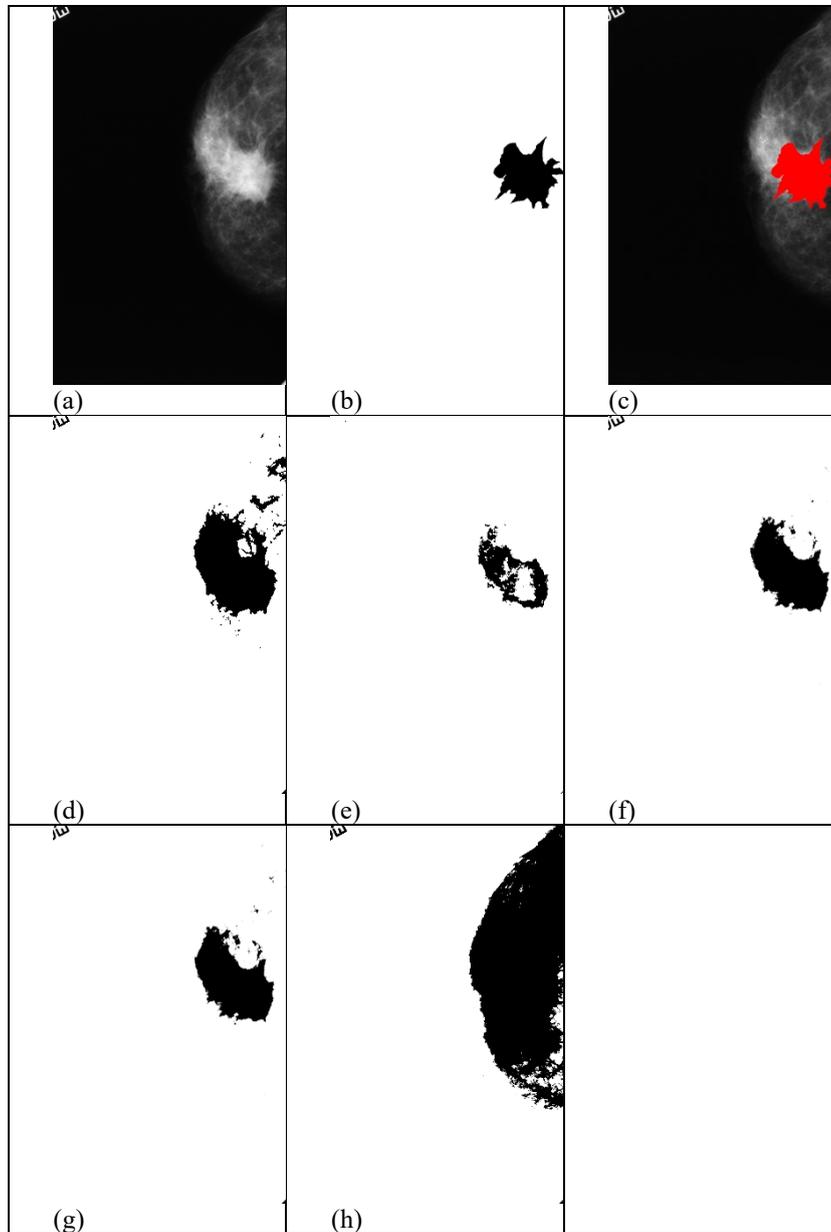
**Tabla 3.** Medida de F promedio en algoritmos de segmentación

Algoritmo	F media
GA	91.697
<b>IWO</b>	<b>93.946</b>
NBA	90.549
OTSU	83.689
PSO	76.414

Con respecto a la Medida de F podemos determinar que tres de los cuatro algoritmos presentan resultados bastante competentes e incluso mejores que el algoritmo de Otsu. Este resultado es relevante por el hecho de tener nuevos algoritmos para realizar segmentación de imágenes de una forma no convencional.

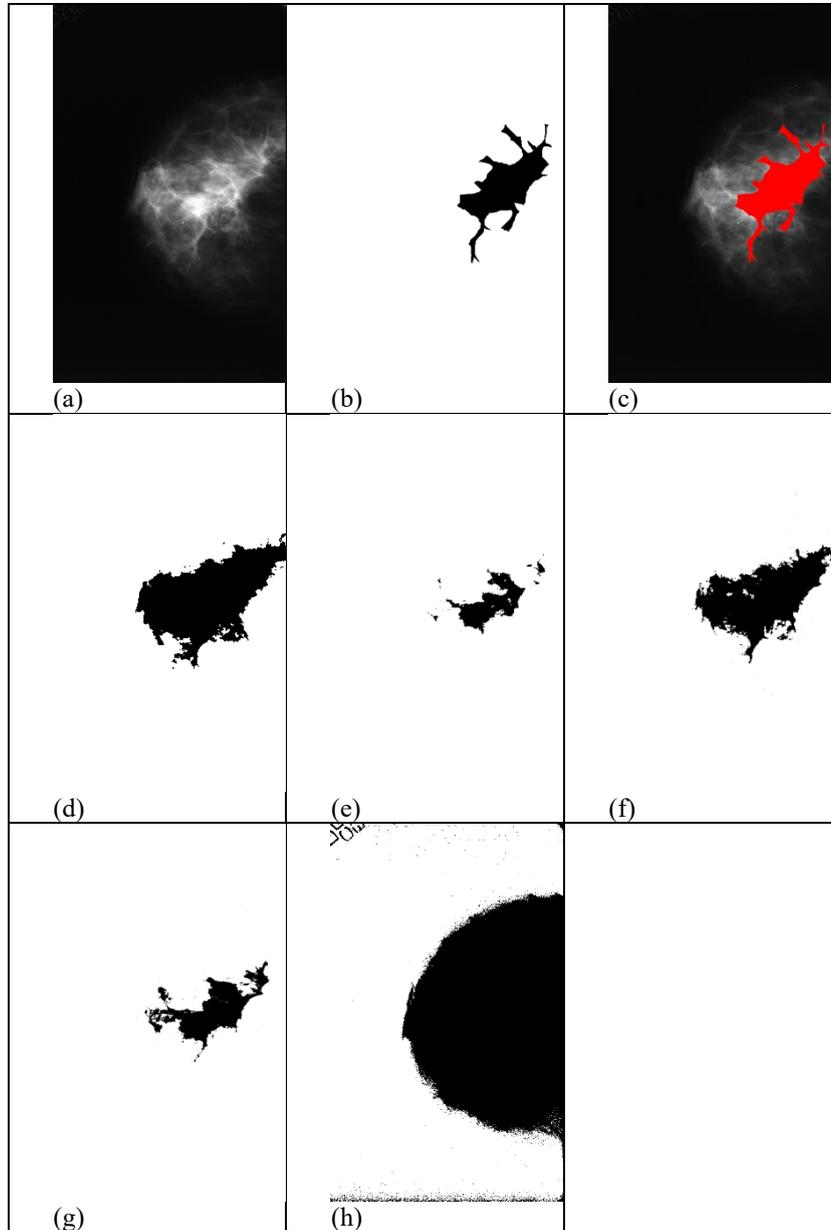
El ajuste de parámetros podría beneficiar al algoritmo PSO ya que no se realizó un ajuste de los mismos.

De forma similar, las Figuras 3, 4 y 5 presentan ejemplos de imágenes tomadas aleatoriamente para ejemplificar los algoritmos y los resultados de cada algoritmo respectivamente.



**Fig.3.** Comparación 1 de los algoritmos de segmentación propuestos (a) Imagen original, (b) Región de interés segmentada por el experto radiólogo, (c) Superposición de la ROI sobre la imagen original, (d) Imagen segmentada por el método Otsu, (e) Imagen segmentada por el método IWO, (f) Imagen segmentada por el método GA, (g) Imagen segmentada por el método NBA, (h) Imagen segmentada por el método PSO

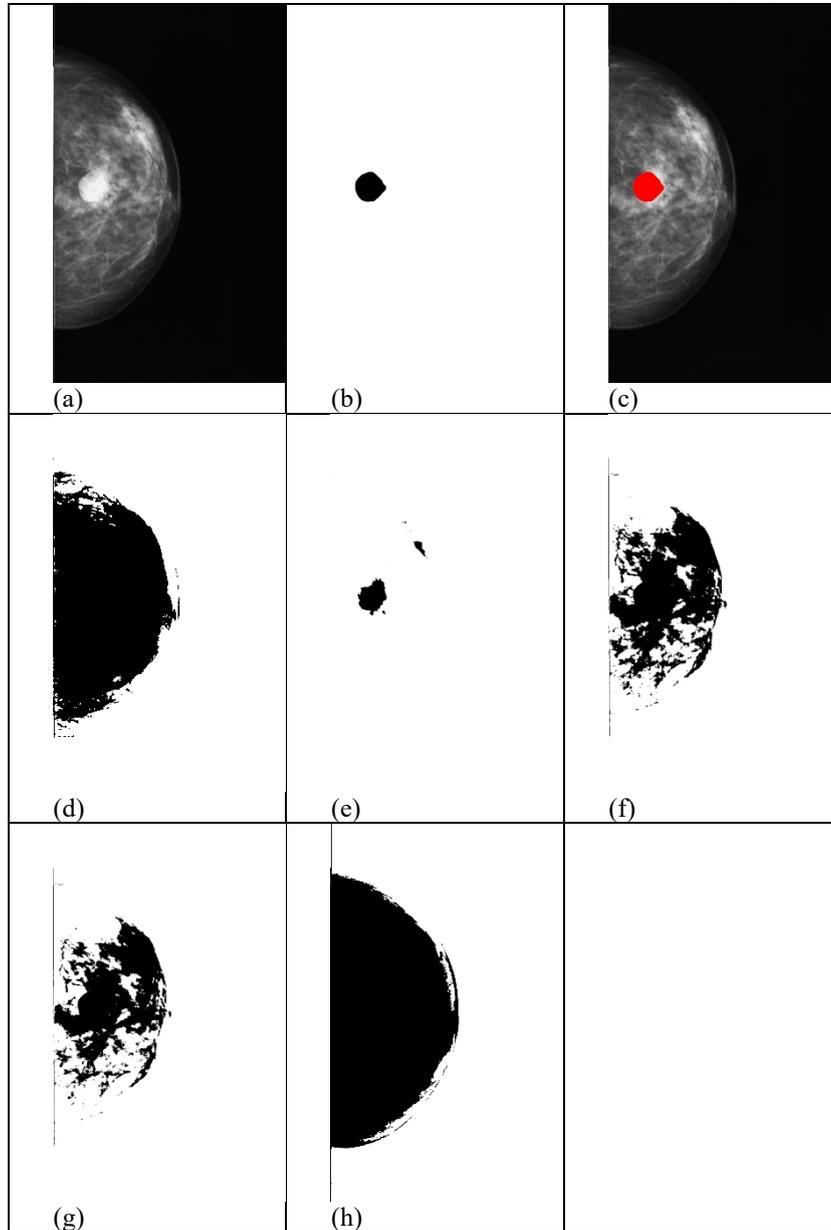
En la figura 3 apreciamos una primera comparación de los algoritmos donde visualmente PSO queda muy alejado de la imagen real de la segmentación. Con respecto a los demás algoritmos, las regiones definidas son muy parecidas, sin embargo algunos algoritmos segmentaron incluso las letras de la esquina superior izquierda. Por otro lado el algoritmo IWO no solo no segmentó dichas letras sino que también obtuvo una región muy similar a la región deseada.



**Fig.4.** Comparación 2 de los algoritmos de segmentación propuestos (a) Imagen original, (b) Región de interés segmentada por el experto radiólogo, (c) Superposición de la ROI sobre la imagen original, (d) Imagen segmentada por el método Otsu, (e) Imagen segmentada por el método IWO, (f) Imagen segmentada por el método GA, (g) Imagen segmentada por el método NBA, (h) Imagen segmentada por el método PSO

En esta segunda comparación podemos recalcar la mala segmentación de PSO ya que como se había mostrado en las Tablas 2 y 3, este algoritmo obtuvo los peores resultados y en esta comparación podemos darnos una idea del porqué. Los demás algoritmos obtuvieron regiones similares a la región buscada presentada en la figura 4 (b).

A pesar de tener regiones similares, algunos algoritmos como Otsu y GA asignaron incorrectamente más píxeles a la región de interés.



**Fig.5.** Comparación 3 de los algoritmos de segmentación propuestos (a) Imagen original, (b) Región de interés segmentada por el experto radiólogo, (c) Superposición de la ROI sobre la imagen original, (d) Imagen segmentada por el método Otsu, (e) Imagen segmentada por el método IWO, (f) Imagen segmentada por el método GA, (g) Imagen segmentada por el método NBA, (h) Imagen segmentada por el método PSO

Finalmente presentamos una última comparación donde se aprecia visualmente el desempeño de los algoritmos y la razón por la cual IWO obtuvo los mejores desempeños incluso comparado con un algoritmo clásico de segmentación y otras metaheurísticas.

En la figura 5 podemos observar que el algoritmo IWO obtuvo una región de interés muy parecida a la región deseada, por lo que si promediamos los resultados de cada algoritmo en las 359 imágenes, eventualmente obtendremos los desempeños presentados en las tablas correspondientes.

## **6. Conclusiones y trabajo a futuro**

En este trabajo se presentaron metaheurísticas aplicadas en segmentación las cuales mostraron mejores desempeños en las dos medidas que se utilizaron. Es interesante comprender que Otsu es un método de umbralado bastante común y utilizado por su facilidad, sin embargo fue posible superar los desempeños obtenidos con dicho método utilizando las metaheurísticas presentadas en este trabajo.

De igual forma se hace uso de las metaheurísticas como algoritmos de segmentación, esto debido a que en su mayoría todas las metaheurísticas son principalmente diseñadas para optimización.

El algoritmo IWO obtuvo los mejores desempeños demostrando ser bastante útil para realizar la segmentación de las imágenes aun considerando que no se realizaron pruebas variando los parámetros.

Por otro lado, no se realizó ningún pre o post procesamiento sobre la imagen mamográfica, esto puede mejorar el contraste y eliminar píxeles de ruido permitiendo obtener una mejor imagen segmentada delimitando de mejor manera la región de interés.

Se contempla como trabajo a futuro la posibilidad de asignar automáticamente los parámetros de las metaheurísticas utilizadas. De igual forma realizar un estudio del impacto de estos parámetros en la segmentación de estos algoritmos.

Esto debido a que el algoritmo PSO obtuvo bajos desempeños, sin embargo un ajuste de sus parámetros podría resultar en lograr una mejor segmentación con este algoritmo. Tomando como ejemplo el algoritmo PSO, es posible aumentar la cantidad de partículas para encontrar mejores soluciones, sin embargo esto también incrementa el tiempo computacional del algoritmo por lo que es necesario determinar un valor óptimo para estos parámetros.

## **Agradecimientos**

Los autores desean agradecer al Instituto Politécnico Nacional (Secretaría Académica, Comisión de Operación y Fomento de Actividades Académicas, Secretaría de Investigación y Posgrado, Centro de Investigación en Computación y Centro de Innovación y Desarrollo Tecnológico en Cómputo), el Consejo Nacional de Ciencia y Tecnología (Conacyt) y Sistema Nacional de Investigadores en México por su apoyo económico para desarrollar este trabajo.

## Referencias

- 1 Ferlay, J., Héry, C., Autier, P. & Sankaranarayanan, R. in Breast cancer epidemiology 1–19 (Springer, 2010).
- 2 Pham, D. L., Xu, C. & Prince, J. L. Current methods in medical image segmentation 1. *Annu. Rev. Biomed. Eng.* 2, 315–337 (2000).
- 3 Dey, V., Zhang, Y. & Zhong, M. A review on image segmentation techniques with remote sensing perspective. (na, 2010).
- 4 Binitha, S., Sathya, S. S. & others. A survey of bio inspired optimization algorithms. *Int. J. Soft Comput. Eng.* 2, 137–151 (2012).
- 5 Fu, K.-S. & Mui, J. K. A survey on image segmentation. *Pattern Recognit.* 13, 3–16 (1981).
- 6 Dollár, P. & Zitnick, C. L. Fast edge detection using structured forests. *IEEE Trans. Pattern Anal. Mach. Intell.* 37, 1558–1570 (2015).
- 7 Malik, J., Dahiya, R., Girdhar, D. & Sainarayanan, G. Finger knuckle print authentication using Canny edge detection method. *Int. J. Signal Imaging Syst. Eng.* 9, 333–341 (2016).
- 8 Mencattini, A., Salmeri, M., Lojaco, R., Frigerio, M. & Caselli, F. Mammographic images enhancement and denoising for breast cancer detection using dyadic wavelet processing. *IEEE Trans. Instrum. Meas.* 57, 1422–1430 (2008).
- 9 Yao, H., Duan, Q., Li, D. & Wang, J. An improved K-means clustering algorithm for fish image segmentation. *Math. Comput. Model.* 58, 790–798 (2013).
- 10 Salem, N. M. Segmentation of white blood cells from microscopic images using K-means clustering. in *Radio Science Conference (NRSC), 2014 31st National* 371–376 (2014).
- 11 Otsu, N. A threshold selection method from gray-level histograms. *Automatica* 11, 23–27 (1975).
- 12 Lu, C., Zhu, P. & Cao, Y. The segmentation algorithm of improvement a two-dimensional Otsu and application research. in *Software technology and engineering (ICSTE), 2010 2nd international conference on* 1, V1--76 (2010).
- 13 Yan, W. G., Wang, C. J. & Guo, J. One extended OTSU flame image recognition method using RGBL and stripe segmentation. in *Applied Mechanics and Materials* 121, 2141–2145 (2012).
- 14 Chen, Q. et al. Modified two-dimensional Otsu image segmentation algorithm and fast realisation. *IET Image Process.* 6, 426–433 (2012).
- 15 Pal, N. R. & Pal, S. K. A review on image segmentation techniques. *Pattern Recognit.* 26, 1277–1294 (1993).
- 16 Banzhaf, W., Nordin, P., Keller, R. E. & Francone, F. D. Genetic programming: an introduction. 1, (Morgan Kaufmann San Francisco, 1998).
- 17 Yang, X.-S. & Hossein Gandomi, A. Bat algorithm: a novel approach for global engineering optimization. *Eng. Comput.* 29, 464–483 (2012).
- 18 Meng, X.-B., Gao, X. Z., Liu, Y. & Zhang, H. A novel bat algorithm with habitat selection and Doppler effect in echoes for optimization. *Expert Syst. Appl.* 42, 6350–6364 (2015).
- 19 Mehrabian, A. R. & Lucas, C. A novel numerical optimization algorithm inspired from weed colonization. *Ecol. Inform.* 1, 355–366 (2006).
- 20 Kennedy, J. & Eberhart, R. Particle swarm optimization (PSO). in *Proc. IEEE*

International Conference on Neural Networks, Perth, Australia 1942–1948 (1995).

**21** Moura, D. C. & López, M. A. G. An evaluation of image descriptors combined with clinical data for breast cancer diagnosis. *Int. J. Comput. Assist. Radiol. Surg.* 8, 561–574 (2013).

**22** Dunn, J. C. A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters. (1973).

**23** Zhao, Y. & Karypis, G. Criterion functions for document clustering: Experiments and analysis. (2001).

**24** Wei, Y. et al. Ts2c: Tight box mining with surrounding segmentation context for weakly supervised object detection. in *Proceedings of the European Conference on Computer Vision (ECCV)* 434–450 (2018).

**25** Fung, B. C. M., Wang, K. & Ester, M. Hierarchical document clustering using frequent itemsets. in *Proceedings of the 2003 SIAM international conference on data mining* 59–70 (2003).

**26** Pan, X. et al. Graph-based RGB-D Image Segmentation Using Color-directional-region Merging. in *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* 2417–2421 (2019).



# e.ciencia

Encuentro Latinoamericano

2 0 1 9



## BELLA

Building the Europe Link to Latin America

