

SCIRE

Representación y organización del conocimiento

SCIRE

Representación y organización del conocimiento

Vol. 29, n.º 2, julio-diciembre 2023

ISSN 1135-3716

ISSN (e) 2340-7042

Scire:
knowledge representation and organization
Vol. 29, n. 2, July-December 2023

Ibersid:
Red de Investigación
en Sistemas de Información
y Documentación

© 2023 Los autores y autoras conservan sus derechos de autor, aunque ceden a la revista de forma no exclusiva los derechos de explotación (reproducción, distribución, comunicación pública y transformación) y garantizan a esta el derecho de primera publicación de su trabajo, el cual estará simultáneamente sujeto a la licencia CC BY-NC-ND. Los autores aceptan la responsabilidad legal de cumplir plenamente con los códigos éticos y leyes apropiadas, y de obtener todos los permisos de derecho de autor debidos. Se permite y se anima a los autores y autoras a difundir electrónicamente la versión editorial (versión publicada por la editorial) en la página web personal del autor y en el repositorio de la institución a la que pertenece.

ISSN: 1135-3716 = Scire (Zaragoza)

ISSN (e): 2340-7042

Depósito legal: Z. 1.790 — 1995

Edita: Ibersid® con la colaboración de Prensas de la Universidad de Zaragoza

Imprime:

Servicio de Publicaciones. Universidad de Zaragoza.

Edificio de Ciencias Geológicas, C/ Pedro Cerbuna, 12.

50009 Zaragoza, España. Tel.: 976 761 330. Fax: 976 761 063.

Scire

representación y organización
del conocimiento

Alcance y objetivos

Scire: representación y Organización del Conocimiento es una publicación semestral de carácter interdisciplinar sobre la representación, normalización, tratamiento, recuperación y comunicación de la información y el conocimiento.

Difusión

Scire tiene difusión internacional. Agradecemos la inclusión en los siguientes servicios de referencia: Scopus, ESCI, Information Science Abstracts, Information Services in Physics, Electronics and Computing, Library and Information Science Abstracts, Sociological Abstracts, ERIH Plus, Knowledge Organization Literature, Base de Datos ISOC y Catálogo Latindex.

Instrucciones para los autores y procedimiento de evaluación

La última versión de las instrucciones para presentación de trabajos y del procedimiento de evaluación editorial están disponibles en: <https://www.iberid.eu/ojs/index.php/scire/about/submissions>

Agradecimientos

Agradecemos el apoyo del Departamento de Ciencia, Universidad y Sociedad del Conocimiento del Gobierno de Aragón con su subvención a grupos de investigación T59_23R: IAAA - Sistemas de Información Avanzados, al Vicerrectorado de Investigación y a la Facultad de Filosofía y Letras de la Universidad de Zaragoza.

Redacción, distribución y canje

Revista Scire
Departamento de Ciencias de la Documentación e Historia de la Ciencia
Facultad de Filosofía y Letras
Universidad de Zaragoza
C/ Pedro Cerbuna 12,
E-50.009 Zaragoza (Spain)

Tfno: int + 34 976 762239. Fax: 34 976761506.
E-mail: mailto:scire@unizar.es

Suscripciones y números sueltos

Suscripción anual: 30 €. Número suelto: 20 €. (IVA inc.)

Scire

knowledge organization
and representation

Scope and aims

Scire: Representación y Organización del Conocimiento is an interdisciplinary journal published twice a year that is devoted to the representation, standardization, treatment, retrieval and communication of information and knowledge.

Dissemination

Scire has international distribution. We acknowledge its inclusion in the following reference services: Scopus, ESCI, Information Science Abstracts, Information Services in Physics, Electronics and Computing, Library and Information Science Abstracts, Sociological Abstracts, ERIH Plus, Knowledge Organization Literature, Base de Datos ISOC and Catálogo Latindex.

Instructions for authors and evaluation process

The last version of the instructions for authors and assessment process is available at: <https://www.iberid.eu/ojs/index.php/scire/about/submissions>

Acknowledgments

We acknowledge the help of the Department of Science, University and Knowledge Society of the Government of Aragón (grant to research group T59_23R: IAAA – Advanced Information Systems), and of the Research Vice Rectorate and the Faculty of Philosophy and Arts of the University of Zaragoza.

Contact address

Revista Scire
Departamento de Ciencias de la Documentación e Historia de la Ciencia
Facultad de Filosofía y Letras
Universidad de Zaragoza
C/ Pedro Cerbuna 12,
E-50.009 Zaragoza (Spain)

Tel.: int + 34 976 762239. Fax: 34 976761506.
E-mail: scire@unizar.es

Subscriptions

Annual subscription: 30 €. Issue: 20 €. (VAT included)

Editor

Francisco Javier García Marco, Univ. de Zaragoza. E-mail: jgarcia@unizar.es

Consejo de redacción / Editorial council

Mario Guido Barité Roqueta,
Universidad de La República, Uruguay

José Augusto Chaves Guimarães,
Universidade Estadual Paulista, Brasil

João Batista Ernesto Moraes,
Universidade Estadual Paulista, Brasil

Francisco Javier García Marco,
Universidad de Zaragoza, España

Daniel Martínez Ávila,
Universidad de León, España

Francisco Javier Martínez Mendez,
Universidad de Murcia, España

Álvaro Quijano Solís,
Colegio de México, México

Consejo científico / Scientific council

Isidro Aguillo Caño, IPP-CSIC, España
Tomás Baiget, EPI S. A., España
José Luis Bonal Zazo, Univ. de Extremadura, España
Mercedes Caridad Sebastián,
Universidad Carlos III de Madrid, España
Alberto Carreras Gargallo,
Universidad de Zaragoza, España
Constança Espelt Busquets,
Universidad de Barcelona, España
Juan Carlos Fernández Molina,
Univ. de Granada, España
María Eulalia Fuentes Pujol, Universidad Autónoma de Barcelona, España
Fernando Galindo Ayuda,
Universidad de Zaragoza, España
Blanca Gil Urdiciain, Universidad Complutense de Madrid, España
Isidoro Gil Leiva,
Universidad de Murcia, España
Alan Gilchrist, Cura Consortium, Reino Unido
Vicente Pablo Guerrero Bote, Universidad de Extremadura, España

Víctor Herrero Solana,
Univ. de Granada, España
José María Izquierdo Arroyo,
Universidad de Murcia, España
María Pilar Lasala Calleja,
Universidad de Zaragoza, España
Alfonso López Yepes, Universidad Complutense de Madrid, España
José López Yepes, Universidad Complutense de Madrid, España
Pedro Marijuán Fernández,
Universidad de Zaragoza, España
Bonifacio Martín del Brío,
Universidad de Zaragoza, España
José Antonio Moreiro González,
Universidad Carlos III de Madrid, España
Purificación Moscoso Castro,
Universidad de Alcalá, España
Félix Moya Anegón,
Universidad de Granada, España
María del Carmen Negrete Gutiérrez,
Universidad Autónoma de México
Catalina Naumis Peña,
Universidad Autónoma de México

José Luis Otaí, Universidad Jaime I de Castellón, España
Manuel José Pedraza Gracia,
Universidad de Zaragoza, España
María Pinto Molina,
Universidad de Granada, España
Gloria Ponjuán Dante,
Universidad de La Habana, Cuba
Blanca Rodríguez Bravo,
Universidad de León, España
José Vicente Rodríguez Muñoz,
Universidad de Murcia, España
Adelaida Román Román,
CINDOC (Madrid), España
Juan Ros García,
Universidad de Murcia, España
Francisco José Ruiz de Mendoza Ibáñez,
Universidad de La Rioja, España
Félix Sagredo Fernández,
Universidad Complutense de Madrid, España
Elías Sanz Casado,
Universidad Carlos III de Madrid, España
Carlos Serrano Cinca,
Universidad de Zaragoza, España

Revisores externos del número / External reviewers in this issue

Agradecemos la colaboración altruista y desinteresada a los Dres. Luisa Alvite Díez, Lourdes Castelló Cogollos, Ricardo Eito Brun, Francisca Hernández Carrascal, Mónica Izquierdo Alonso, Luis Martínez Uribe, Walter Moreira, María Antonia Ovalle Perandones, Jesús Robledano Arillo y Juan Miguel Sánchez Vigil.

Candidaturas al consejo científico

Se aceptan candidaturas al consejo científico de especialistas del área de Biblioteconomía y Documentación y de otras disciplinas relacionadas (Informática, Ciencias Sociales, Lingüística, Filosofía, Psicología, etc.) con experiencia profesional e investigadora demostrada. En el sistema público de investigación español, suele ser equivalente al doctorado y dos sexenios de investigación o méritos equivalentes.

Scientific council membership policy

Candidatures of researchers from LIS and other related disciplines (Computer Science, Social Sciences, Linguistics, Philosophy, Psychology, etc.) with demonstrated professional and research experience are welcomed. In the Spanish public research system, for example, this usually means having a doctorate and two scientific productivity sexennia or equivalent outputs.

Tabla de contenidos en español

Table of contents in Spanish

Tabla de contenidos en español 9

Tabla de contenidos en inglés 11

Artículos

Ontologías terminológicas en la organización del conocimiento: una aproximación bibliométrica

Adriana SUÁREZ SÁNCHEZ..... 13

Fundamentos lingüísticos de la indización: una revisión

Alexandre Robson MARTINES
Carlos Cândido ALMEIDA..... 25

Implementación de los repositorios de datos de investigación en las universidades públicas españolas: estado de la cuestión

Francisco Javier MARTÍNEZ MÉNDEZ
Ana Alice BAPTISTA
Rosana LÓPEZ CARREÑO
Ángel María DELGADO VÁZQUEZ..... 39

Análisis reticular de las relaciones de colaboración de las empresas de desechos tecnológicos de Panamá

Ángel ZAZO
Saúl ARDINES
Amarilis DE LEÓN
María Luisa PAZ
Virginia JUÁREZ
Isaac PERDOMO..... 51

Inteligencia artificial y decisiones judiciales: crónica de una transformación anunciada

Carolina SANCHIS CRESPO..... 65

Un modelo curatorial para las colecciones de patrimonio digital como datos abiertos enlazados y grafos con nombre

Carlos H. MARCONDES 85

Índice de autores..... 96

Índice de materias en español 96

Índice de materias en inglés 96

Tabla de contenidos en inglés

Table of contents in English

Table of contents in Spanish..... 9
Table of contents in English 11

Articles

Terminological ontologies in knowledge organization: a bibliometric approach
Adriana SUÁREZ SÁNCHEZ 13

Linguistic fundamentals of indexing: a review
Martines, Alexandre Robson, Almeida, Carlos Cândido..... 25

Implementation of research data repositories in Spanish public universities: state of the art
Francisco Javier MARTÍNEZ MÉNDEZ
Ana Alice BAPTISTA
Rosana LÓPEZ CARREÑO
Ángel María DELGADO VÁZQUEZ..... 39

Network analysis of collaborative relationships of WEEE companies in Panama
Ángel ZAZO
Saúl ARDINES
Amarilis DE LEÓN
María Luisa PAZ
Virginia JUÁREZ
Isaac PERDOMO..... 51

Artificial Intelligence and judicial decisions: chronicle of an announced transformation
Carolina SANCHIS CRESPO..... 65

A curatorial model for digital heritage collections as linked open data and named graphs
Carlos H. MARCONDES 85

Author index..... 96
Subject index in Spanish..... 96
Subject index in English..... 96

Ontologías terminológicas en la organización del conocimiento: una aproximación bibliométrica

Terminological ontologies in knowledge organization: a bibliometric approach

Adriana SUÁREZ SÁNCHEZ

Instituto de Investigaciones Bibliotecológicas y de la Información, Universidad Nacional Autónoma de México, Circuito Interior de Ciudad Universitaria s/n Pisos 11, 12 y 13 de la Torre II de Humanidades, Ciudad Universitaria, Alcaldía Coyoacán, Ciudad de México, México. asuarez@iibi.unam.mx

Resumen

Las ontologías terminológicas son un sistema de organización del conocimiento que en los últimos años ha ganado gran popularidad debido a sus ventajas en la representación y estructuración de dominios en el contexto digital. En la literatura se han tratado sus aspectos teóricos y sus aplicaciones, pero no se cuenta con suficientes estudios bibliométricos. En el presente artículo se analiza la producción científica sobre ontologías terminológicas desde una aproximación bibliométrica. Se estudia la producción científica sobre el tema registrada en Web of Science, Scopus, WorldCat, Dialnet, Redalyc y Google Académico. Se analiza: el rango temporal de la literatura, su crecimiento diacrónico, la tipología de los documentos publicados, los idiomas de publicación, las publicaciones periódicas con mayores trabajos, la productividad de los autores, el grado de colaboración y el rango de citación. Se encontraron 216 documentos producidos por 420 autores en un rango temporal de 1997 a 2022. Los documentos publicados corresponden a libros, capítulos de libros, artículos en revistas académicas, trabajos en actas de congresos y tesis. El idioma de publicación preponderante es inglés. El núcleo de productores se forma de 13 autores. La producción sobre el tema muestra un pico de crecimiento entre 2007 y 2016, actualmente registra una tendencia alta de producción y en la literatura prevalece un alto grado de colaboración.

Palabras clave: Ontologías terminológicas. Organización del conocimiento. Bibliometría. Producción científica.

1. Introducción

Las ontologías son “una explicitación de una conceptualización compartida” (Gruber, 1995, p. 908) o bien “un marco conceptual particular en el nivel del conocimiento” (Guarino, 1995, p. 633) que se concreta en símbolos y es utilizado con fines prácticos de gestión del conocimiento.

Por mucho tiempo, tales sistemas fueron asunto exclusivo de las ciencias computacionales, en tareas de modelado de conocimiento en agentes computacionales; sin embargo, al paso del

Abstract

Terminological ontologies are a knowledge organization system that has gained great popularity in recent years due to its advantages in the representation and structuring of domains in the digital context. Theoretical aspects and applications have been discussed in the literature, but there are not enough bibliometric studies. This article analyses the scientific production on terminological ontologies from a bibliometric approach. The scientific production on the subject registered in Web of Science, Scopus, WorldCat, Dialnet, Redalyc and Google Scholar is studied. The following aspects are analysed: the temporal range of the literature, its diachronic growth, the typology of the published documents, the languages of publication, the journals with the largest number of papers, the productivity of the authors, the degree of collaboration and the citation rank. We found 216 publications by 420 authors over a period from 1997 to 2022. The published documents correspond to books, book chapters, articles in academic journals, papers in conference proceedings and dissertations. The predominant language of publication is English. The core group of producers consists of 13 authors. The production on the subject shows a growth peak between 2007 and 2016, currently shows a high production trend and a high degree of collaboration prevails in the literature.

Keywords: Terminological ontologies. Knowledge organization. Bibliometrics. Scientific research.

tiempo se expandieron hacia diversos campos de saber que organizan el conocimiento, la información y los recursos de información, tal es el caso de la bibliotecología.

García Marco (2008) señala que la expansión del término hacia la bibliotecología constituye la punta del iceberg de un proceso de reconfiguración disciplinar, como resultado de la integración de diversas ciencias en el marco del desarrollo tecnológico info-comunicacional. En el ámbito de la representación, organización y recuperación

de la información y el conocimiento, las ontologías son un sistema de organización que establece el vocabulario de un campo de saber, mediante un conjunto de conceptos, atributos y relaciones entre tales términos. A grandes rasgos, las ontologías pueden considerarse un sistema de organización del conocimiento (Hodge, 2000) con una sintaxis comprensible para los ordenadores (Pedraza, Codina y Rovira, 2007).

Como sistemas de organización del conocimiento, las ontologías cuentan con un gran abanico de funciones (vocabulario controlado, modelado de procesos, soporte de algoritmos para agentes artificiales, etc.) que se asocian con una amplia tipología de los instrumentos. Como resultado, entre los diversos tipos de ontologías persiste una amplia categorización que engloba: ontologías de dominio, ontologías de tarea, ontologías terminológicas, ontologías ligeras, ontologías pesadas, entre otras.

Las ontologías terminológicas, de especial interés en esta investigación, derivan de la tipología propuesta por Van Heijst, Schreiber y Wielinga (1997) quienes hablan de ontologías terminológicas y ontologías modeladoras. Las ontologías terminológicas especifican los términos que son usados en un dominio mientras que las modeladoras se centran en las simbolizaciones del conocimiento y son más ricas en relaciones axiomatizadas. Lacasta, Nogueras-Iso y Zarazaga-Soria (2010, p. 24) señalan que: “dos grandes tipos de modelos de ontologías han sido estudiadas: por un lado, ontologías terminológicas [...] por otro lado, ontologías axiomatizadas o formales”.

Las ontologías modeladoras o axiomáticas son ideales en la representación de fragmentos de conocimiento en agentes computacionales, mientras que las ontologías terminológicas, con funciones de control, explicitación, direccionamiento o vinculación léxica son empleadas con fines terminológicos y como vocabularios controlados en la organización del conocimiento y la indexación de recursos de información. Tenemos entonces que las ontologías con fines terminológicos son teorías de dominios que, mediante el uso del lenguaje, explicitan las entidades de un área de saber, las organizan en un esquema jerárquico y determinan sus propiedades y relaciones, proporcionando un vocabulario y vínculos entre elementos denominativos.

En las últimas dos décadas, las ontologías terminológicas han acaparado la atención de los bibliotecólogos, en vista de sus características como un vocabulario controlado y una estructura categorizadora que las agrega al amplio abanico de los sistemas de organización del conocimiento (SOC) (Hodge, 2000) entre los que es posible

mencionar encabezamientos de materia, tesauros, taxonomías, mapas tópicos, ontologías, entre otros. Aunado a ello, se vislumbran como un SOC con gran potencial para el modelado temático-terminológico de campos de conocimiento (endometriosis, tumores cerebrales, plantas suculentas, canciones folclóricas búlgaras, Colección Obras Maestras del Museo del Prado, etc.) y la organización de recursos de información relacionados.

A partir de su relevancia en la organización del conocimiento, la producción científica sobre ontologías terminológicas ha sido un asunto relevante en las últimas dos décadas; no obstante, los estudios bibliométricos sobre el tema son escasos.

García Marco (2008) establece que los artículos sobre ontologías eran muy raros hasta 1992. Luego, se observó un aumento en la producción científica durante el periodo de 1990 a 1998. Por último, hacia principios de milenio hubo una consolidación del tema, especialmente en el contexto de la web semántica y el problema terminológico-conceptual en ambientes informatizados.

Un segundo trabajo que aportó información respecto a la producción científica sobre ontologías fue desarrollado por Zhu, Kong, Hong, Li y He (2015). En su artículo estudiaron la literatura registrada en Web of Science y determinaron que la investigación ontológica ha entrado en una fase de auge, caracterizada por un alto grado de colaboración entre autores.

En 2017, Netto y Lima estudiaron la producción bibliométrica sobre modelos y tecnologías para la visualización de las ontologías. Detectaron un nicho de investigación reciente, con un rango de publicación entre 2007 a 2017. Las autoras concluyeron que muchas aristas de las ontologías se encuentran en un estado de auge de investigación y que la producción científica crece día a día, por lo que es necesario llevar a cabo estudios bibliométricos que permitan analizar el dominio, los autores, las fuentes, etc.

A partir del estado del arte del tema, revisado en los párrafos anteriores, es posible recapitular dos aspectos: las múltiples aristas de estudio de las ontologías (epistemología, estructura, tecnologías, formas de visualización, entre otros) y los pocos estudios bibliométricos, que derivan en la necesidad de analizar su producción científica.

Consecuentemente, surgen algunas preguntas de investigación:

- ¿Cuál es la tendencia de la producción científica sobre ontologías terminológicas en funciones de organización del conocimiento?
- ¿A través de qué tipo de documentos se distribuye dicha producción científica?

- ¿En qué idioma se produce la literatura sobre ontologías terminológicas?
- ¿Quiénes son los autores con mayor producción científica?
- ¿La colaboración se observa como una característica de dicha literatura?
- ¿Qué prospectiva de crecimiento presenta el tema en los tiempos futuros?

En aras de dar respuesta a las preguntas antes esbozadas, el presente artículo tiene por objetivo: analizar desde una perspectiva bibliométrica la producción científica sobre ontologías terminológicas en funciones de organización del conocimiento. Se enfatiza la relevancia del trabajo a partir de la trascendencia que las ontologías terminológicas poseen en la organización terminológica-temática de dominios de conocimiento, su uso como lenguajes de indización y su potencial en el arreglo de recursos de información albergados en el espacio web.

2. Metodología

Para lograr el objetivo establecido, se siguió una metodología descriptiva-cuantitativa, fundamentada en la técnica de análisis bibliométrico. Dicha técnica de investigación implica “el uso de métodos matemáticos y estadísticos para analizar el desarrollo histórico de un cuerpo específico de literatura, especialmente su autoría, publicación y uso” (Reitz, 2007).

Los estudios bibliométricos son un método nuclear en las ciencias de la información y consideran dos enfoques: el mapeo gráfico y el análisis de funcionamiento. Entre ellos, la presente investigación hace uso del análisis de funcionamiento, un método que analiza aspectos generales de la producción científica (idioma, autores, tipología de la publicación) así como indicadores de productividad y citación (Aparicio, Iturralde y Maceda, 2019).

Por producción científica se entiende el conjunto de obras (libros, artículos, capítulos de artículo, tesis y trabajos compendiados en actas de congresos) resultantes del trabajo intelectual de la investigación desarrollada en un área de conocimiento (Maletta, 2009). Se optó por medir la producción científica a la luz del análisis bibliométrico porque es claro indicio de la cantidad de investigación producida por un científico, una institución o un país (Spinak, 1996). Piedra Salomón y Martínez Rodríguez (2007, p. 34) agregan que “la productividad científica refleja el resultado de investigaciones científicas traducidas en nuevo conocimiento.”

Una vez establecida la perspectiva de estudio, la investigación comprendió tres etapas: obtención de los registros bibliográficos, sistematización y limpieza de registros bibliográficos, y procesamiento de los datos.

2.1. Obtención de los registros bibliográficos

Para localizar la producción científica sobre ontologías terminológicas en funciones de organización del conocimiento se revisaron las siguientes bases de datos (Tabla I):

Base de datos	Descripción
Scopus	Es una base de datos especializada en información científica, técnica, ciencias médicas y sociales. Incluye artículos de diversos campos de conocimiento. Contiene más de 29 millones de resúmenes de más de 18,500 revistas científicas tanto comerciales como de acceso abierto.
Web of Science	Es una base de datos que contiene información sobre investigación multidisciplinaria, proveniente de más de 3,300 revistas líderes mundiales en las ciencias puras, ciencias sociales, artes y humanidades. Es un referente sobre la producción científica internacional.
Dialnet	Es un portal de difusión de la producción científica hispana. Su base de datos fue creada por la Universidad de la Rioja (España). Incluye índices de revistas científicas y humanísticas de España, Portugal y Latinoamérica. Incorpora también libros, tesis doctorales, congresos, homenajes y otro tipo de documentos.
WorldCat	Es una base de datos sobre recursos albergados en las colecciones de numerosas bibliotecas del mundo. Es gestionada por la cooperativa OCLC.
Google Académico	Es un motor de búsqueda de Google especializado en la búsqueda de bibliografía científico-académica. Indexa bibliotecas, repositorios, bases de datos bibliográficas, etc. Contiene registros de libros, artículos, congresos, informes, tesis. Tesinas y archivos depositados en repositorios.

Tabla I. Bases de datos consultadas

Luego de establecer las bases de datos que se consultarían, se emprendió la búsqueda de los

documentos a partir de un conjunto de descriptores en tres idiomas (inglés, español y portugués). La recuperación de los registros consideró toda la producción científica existente en cada una de las bases de datos hasta diciembre del año 2022.

En la detección de los documentos, se emplearon los siguientes descriptores:

- “ontologías terminológicas” [español]
- “terminological ontologies” [inglés]
- “ontologias terminológicas” [portugués]
- “ontologías” and “términos” [español]
- “ontologies” and “terms” [inglés]
- “ontologias” and “termos” [portugués]
- “ontologías” and “términos” and “organización de conocimiento” [español]
- “ontologies” and “terms” and “knowledge organization” [inglés]
- “ontologias” and “termos” and “organizaçãodo conhecimento” [portugués]

La consulta de tales descriptores se efectuó en los campos de título, temas, resumen y palabras clave, según los campos de búsqueda disponibles en cada una de las bases de datos en las que se detectaron los trabajos (1). A partir del trabajo señalado, se recuperaron 312 registros sobre ontologías terminológicas en funciones de organización del conocimiento que incluyeron tanto trabajos teóricos como aplicados.

2.2. Sistematización y limpieza de registros bibliográficos

Las referencias localizadas en las bases bibliográficas antes enunciadas se exportaron a una base de datos en EndNote X8. Cada documento se revisó cuidadosamente para verificar que tratara sobre ontologías terminológicas en tareas de organización del conocimiento. Se verificó que no hubiese registros duplicados y se normalizaron los nombres de los autores, los títulos de los documentos, los nombres de las revistas y los nombres de los congresos. Tras dichas actividades, de los 312 registros obtenidos se eliminaron 96. Luego de tal acción, el corpus de análisis, que sirvió de base al presente estudio, se constituyó de 216 trabajos sobre ontologías terminológicas publicados desde 1997 hasta 2022 por 420 autores.

2.3. Procesamiento de los datos

En el análisis bibliométrico del corpus sobre ontologías terminológicas se usó el programa SPSS 25.0.

La productividad de los autores se analizó con la aplicación de la ley de Lotka (Urbizagástegui Alvarado, 1999), según el modelo de mínimos cuadrados:

$$Y_x = C \left(\frac{1}{x^n} \right)$$

Donde:

- C = número total de autores con un único documento entre número total de autores
- n = parámetro que se estima de los datos
- x = número de autores con 1, 2, 3... n documentos producidos

Para medir el grado de colaboración, se estableció la teoría del coeficiente de colaboración (Aji-feruke, Burell y Tague, 1988). El CC está basado en la idea de atribución fraccional de productividad en la generación de documentos y puede ser expresado mediante la siguiente ecuación:

$$CC = 1 - \frac{\sum_{j=1}^k \left(\frac{1}{j} \right) f_j}{N}$$

Donde:

- f_j = el número de artículos de investigación con j -autores múltiples publicados en una disciplina durante determinado periodo.
- N = número total de artículos de investigación publicados en una disciplina durante determinado periodo.
- K = el mayor número de autores por artículo en una disciplina.

Para medir el crecimiento de la literatura se empleó la función:

$$C(t) = C(0)^{at}$$

que según la teoría de Egghe y Ravichandra Rao (1992), puede ser desplegada como:

$$C(t) = cg^t \text{ donde } c > 0, g > 1, \text{ y } t \geq 0$$

También en el rubro crecimiento de la producción científica se detectó el periodo de duplicación de la literatura, según la siguiente ecuación en regresión no lineal:

$$n(\log g) = \log 2,0$$

$$n = \frac{\log 2,0}{\log g}$$

Por último, en lo tocante a los datos para determinar los trabajos más citados se consideró el número de referencias a los documentos en Google Académico. Si bien algunos registros bibliográficos fueron recuperados en bases de datos como Scopus o Web of Science, que incluían el

número de citas según el conteo de la base, se optó por el número de citas registrado en Google Académico con fines de homologar la base a partir de las citas directas en una misma fuente de información.

3. Resultados

A partir de la investigación realizada fue posible obtener un panorama general sobre el estado de las ontologías terminológicas desde una perspectiva bibliométrica. En los apartados siguientes se presentan los resultados del estudio en diversos rubros que contemplan el rango temporal de los trabajos, los idiomas de publicación, rangos de citación y otros aspectos bibliométricos.

3.1. Rango temporal de la producción científica

A partir del estudio realizado, se encontró que las ontologías terminológicas en funciones de sistemas de organización del conocimiento son un tema reciente (Gráfico 1).

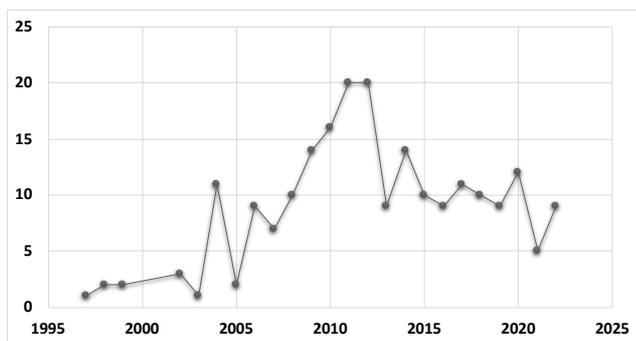


Gráfico 1. Producción científica 1997-2022

En el corpus examinado registran presencia literaria desde 1997. El primer trabajo detectado corresponde a D.M. Pisanelli, A. Gangemi y G. Steve (1997), titulado *WWW-available conceptual integration of medical terminologies: the ONIONS experience*; en tal estudio, los autores examinan la terminología médica a la luz de una arquitectura ontológica.

Las ontologías terminológicas emergieron con gran fuerza en las ciencias computacionales a mediados de la década de 1980, campo en el que fueron empleadas para modelar el mundo en sistemas computacionales, comunicar explicitaciones de conocimiento consensuadas y transferir conocimiento entre agentes mecánicos y humanos, así como entre diversos agentes mecánicos.

Su integración en la bibliotecología, como un sistema de organización del conocimiento se registró hasta una década después. Como señala Suárez Sánchez (2022, p. 39):

Las ontologías, como un sistema de organización del conocimiento, empezaron a tener presencia en la bibliotecología a partir de la década de los noventa. Su introducción se registró en el área de la organización del conocimiento y la información, campo en el que aconteció un traslape teórico-práctico entre las teorías y herramientas de las ciencias computacionales y los postulados de las ciencias documentales.

Según la línea temporal de producción científica (Gráfico 1), el estudio coincide con lo señalado por López Alonso (2000): algunos modelos conceptuales para la representación del conocimiento y la información (tesauros, encabezamientos de materia) cuentan con larga presencia en el ámbito bibliotecológico y son conocidos de antaño; sin embargo, las ontologías, como instrumentos hipermedia y asociadas a ambientes virtuales, son nuevas formas, con apenas unas cuantas décadas de inserción y estudio en la disciplina.

3.2. Crecimiento de la producción científica

Si desglosamos la línea temporal de la producción científica sobre ontologías terminológicas en quinquenios es posible identificar cinco grandes periodos:

- 1997-2001, etapa incipiente
- 2002-2006, etapa de desarrollo
- 2007-2011, etapa de masificación
- 2012-2016, etapa de auge
- 2017-2022, etapa de estabilización.

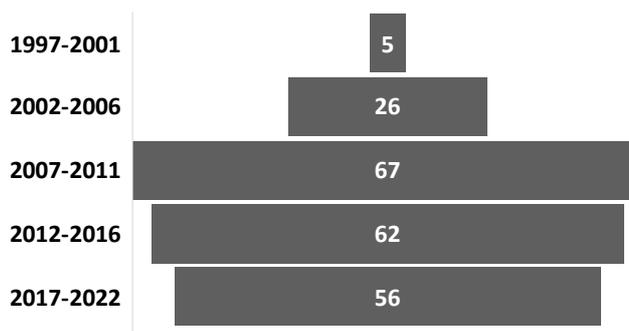


Gráfico 2. Producción científica por periodo

Observamos (Gráfico 2) que antes del año 2000 los estudios sobre ontologías terminológicas eran escasos. El crecimiento de la producción científica empieza a desarrollarse después del año 2002. Luego, en 2007 la literatura sobre el tema registra un momento de masificación que Sánchez-Cuadrado, Morato-Lara, Palacios-Madrid, Llorens-Morillo y Moreiro-González (2007)

mencionan en su artículo *De repente: ¿todos hablamos de ontologías?* En el quinquenio siguiente (2012-2016), la producción científica mantiene una etapa de auge.

Por último, después del año 2017 se advierte un ligero descenso en la producción literaria, tal hecho entraña relación con un efervescente interés por el tema en el rango 2007-2011 y la consolidación del tema en el periodo 2012-2016.

El pico más alto de crecimiento de la literatura se registra en el periodo 2007-2011 y de esos años a la actualidad el tema se ha mantenido como asunto de reflexión y análisis.

Para continuar con la tendencia de la producción científica, se analizó el crecimiento de la literatura, según los preceptos de Egghe y Revichandra Rao (1992).

Si retomamos la fórmula de crecimiento:

$$n(\log g) = \log 2,0$$

$$n = \frac{\log 2,0}{\log g}$$

Los valores con los que contamos para determinar el rango de crecimiento de la producción científica son:

$$c = 5,094$$

$$g = 1,038$$

$$R^2 = 0,230$$

Tomando en cuenta el modelo $C(t) = cg^t$, establecido en la metodología, los valores quedaron como sigue:

$$C(t) = 5,094 * 1,038^t \quad \text{donde } t = 0,038$$

Se determinó que la tasa de crecimiento de la producción científica anual es del 3,8 %. Luego, para estimar la tasa de duplicación se empleó la siguiente ecuación:

$$g * n = 2,0$$

Al reemplazar los valores, se obtuvo:

$$(1,038) n = 2$$

A partir de los logaritmos de ambos lados de la ecuación, se sintetizó que:

$$n(\log 1,038) = \log 2,0:$$

$$n = \frac{\log 2,0}{\log 1,038}$$

$$n = \frac{0,693}{0,037}$$

$$n = 18,7$$

Dado que la unidad de medida de crecimiento son años, se dedujo que la literatura científica sobre ontologías terminológicas en funciones de organización del conocimiento se duplica cada 18 años (2).

3.3. Idiomas de publicación

Respecto al idioma en el que la literatura sobre ontologías terminológicas se publica, el estudio sólo consideró tres lenguas (inglés, portugués y español). El total de 216 documentos, según idioma, se distribuyó del modo siguiente (Gráfico 3).

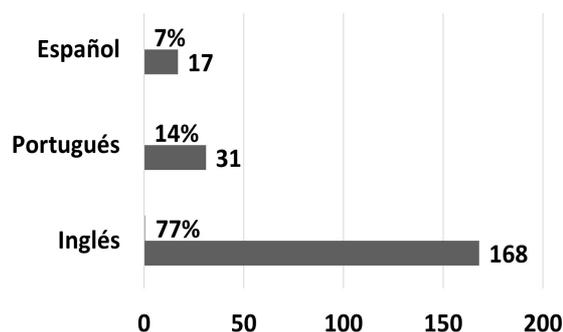


Gráfico 3. Idioma de las publicaciones

El idioma de publicación mayormente empleado es inglés con el 77 %, le siguió portugués con el 14 % y finalmente el idioma español con el 7 %. Tal hecho reafirma la idea de que (Drubin y Kellogg, 2017, p. 1)

[...] el inglés se usa actualmente como el lenguaje de la ciencia [...] esta adopción de un idioma universal o lengua franca ha tenido un efecto extraordinario en la comunicación científica, permitiendo a los investigadores de todo el mundo tener acceso a la vasta producción de artículos y haciendo posible la comunicación entre científicos en cualquier país.

Si bien un número considerable de estudios son presentados en el idioma local del investigador y de su comunidad de lectores más cercana, es recurrente que muchos expertos decidan publicar en idiomas extranjeros. Aunado a ello, en el caso de las ontologías terminológicas fue posible identificar que el idioma dominante guarda relación con el tipo de recurso dominante: actas de congresos, que suelen adherirse a un idioma establecido (inglés) con miras a la comunicación presencial entre los expertos que asisten a la convención y la publicación de las actas del encuentro. En el caso de los congresos, frecuentemente, los expertos se ven obligados a presentar sus comunicaciones orales y escritas en lengua inglesa que, consecuentemente, registra un mayor número de publicaciones.

3.4. Tipología de la producción científica

En cuanto al tipo de recursos en los que la literatura sobre ontologías terminológicas se difunde, se detectó lo siguiente (Gráfico 4):

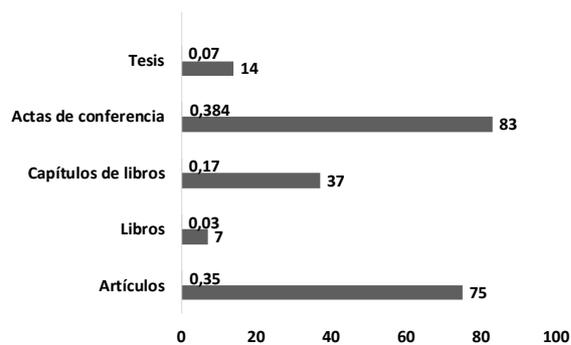


Gráfico 4. Tipo de recursos

Como puede apreciarse, las actas de conferencia, derivadas de los encuentros de la especialidad, generan la mayor producción científica sobre el tema, con 83 casos que corresponden al 38% de los registros analizados. En dicho rubro es importante destacar la relevancia de eventos internacionales como *Proceedings of the Iberoamerican Meeting of Ontological Research*, *CEUR Workshop Proceedings*, *Proceedings of The Semantic Web in Libraries Conference*, *Proceedings of The International Conference on Knowledge Engineering and Ontology Development* y *Advances in Knowledge Organization*.

Tras la fuerte tendencia de publicación sobre ontologías terminológicas en encuentros de la especialidad, las revistas son el segundo recurso con mayor literatura, con el 35% de la publicación. Los capítulos de libros son el tercer tipo de recurso con el 17%, le siguen las tesis con el 7% y los libros con el 3%.

En lo relativo a la presencia de los artículos académicos en títulos de revistas, los 75 artículos detectados se registran en 62 títulos de revistas. Siete títulos de revistas concentran el 25% de los artículos y el 75% de los artículos restantes se distribuyen en 55 títulos de revistas que sólo registran un artículo sobre el tema. Tal fenómeno confirma la idea de que, frecuentemente, un número reducido de publicaciones altamente especializadas reúne un número considerable de la producción científica sobre un tópico, mientras que el resto de las revistas son de frontera, generales o de dispersión (Urbizagástegui Alvarado, 2021).

Las revistas con mayor presencia de artículos sobre ontologías terminológicas son (Tabla 2):

Revista	Artículos
Applied ontology: an interdisciplinary journal of ontological analysis and conceptual modeling	4
Ciência da informação	3
IEEE Transactions on knowledge and data engineering	3
Onomázein: revistas de filología, lingüística y traducción	3
Perspectivas em ciência da informação	2
Revista signos	2
Scire: representación y organización del conocimiento	2

Tabla II. Revistas con mayor producción científica

Entre las publicaciones periódicas destaca la revista *Applied ontology: an interdisciplinary journal of ontological analysis and conceptual modeling* que es la principal fuente de diseminación de investigaciones. Tal publicación, comprende dos grandes rubros: el análisis ontológico y el modelado conceptual. Según el análisis ontológico, integra trabajos sobre epistemología, modelado de dominios y lógica para la estructuración de redes semánticas de información. Según el modelado conceptual, incluye estudios sobre las estructuras cognitivas, categoriales y lingüísticas para modelar el mundo en agente computacionales, así como la terminología que es la base de diversos sistemas de organización del conocimiento (tesauros, taxonomías, ontologías, entre otros).

Entre las revistas con mayor producción sobre ontologías terminológicas destacan dos revistas brasileñas que publican en portugués sobre el tema: *Ciência da informação* y *Perspectivas em ciência da informação*. Aunado a ello, el trabajo más citado en el rubro también está escrito en portugués. A partir de ello, es posible señalar una contribución significativa de la producción científica en lengua portuguesa en la investigación sobre ontologías.

3.5. Rango de citación

Los trabajos más citados fueron otro elemento de análisis en la producción científica del tema. Se encontró que los diez trabajos más citados con sus respectivos autores son los siguientes (Tabla III, en la página siguiente).

El trabajo más citado *Uma visão geral sobre ontologias: pesquisa sobre definições, tipos, aplicações, métodos de avaliação e de construção* (Almeida y Bax, 2003) es una revisión general sobre las ontologías que sólo abunda sobre las ontologías terminológicas como uno de sus tipos. Al

respecto señala que la relevancia de las ontologías terminológicas en el ámbito documental reside en su alcance como sistemas para la organización temática de dominios de conocimiento, la indización de recursos de información en bibliotecas digitales y su utilidad como plantillas conceptuales para la recuperación de la información en la web.

Autores	Título	Citas
Almeida, M. B. y M. P. Bax	Uma visão geral sobre ontologias: pesquisa sobre definições, tipos, aplicações, métodos de avaliação e de construção	352
Borge-Holthoefer, J. y A. Arenas	Semantic networks: structure and dynamics	226
Wang W., P.M. Barnaghi y A. Bargiela	Probabilistic topic models for learning terminological ontologies	142
García Jiménez, A.	Instrumentos de representación del conocimiento: tesauros versus ontologías	123
Nielsen, L.P.	Knowledge dissemination based on terminological ontologies	82
Rani, M., A.K. Dhar y O.P. Vyas	Semi-automatic terminology ontology learning based on topic modeling.	60
Lacasta, J., J. Noguera-Iso y F.J. Zarazaga-Soria	Terminological ontologies: design, management and practical applications	38
Madsen B.N., H.E. Thomsen y C. Vikner	Principles of a system for terminological concept modelling	36
Jiménez-Briones, R. y A. Luzondo Oyon	Building ontological meaning in a lexico-conceptual knowledge base	35
Soler Monreal, C. e I. Gil Leiva	Posibilidades y límites de los tesauros frente a otros sistemas de organización del conocimiento: folksonomías, taxonomías y ontologías	34

Tabla III. Trabajos con mayor número de citas

3.6. Productividad de los autores

En lo tocante a la productividad de los autores, en la investigación se siguieron los preceptos de la Ley de Lotka. Se hizo una comparación de la productividad de los autores considerando un rango estimado y un rango observado. También se realizó un conteo completo de todos los autores y se obtuvo un valor para $c = 0,838$, $n = 3,07$ con un error estándar de 0,081 y con un R^2 de 0.95, es decir, una confiabilidad de 95%.

En tal rubro se obtuvieron los siguientes resultados:

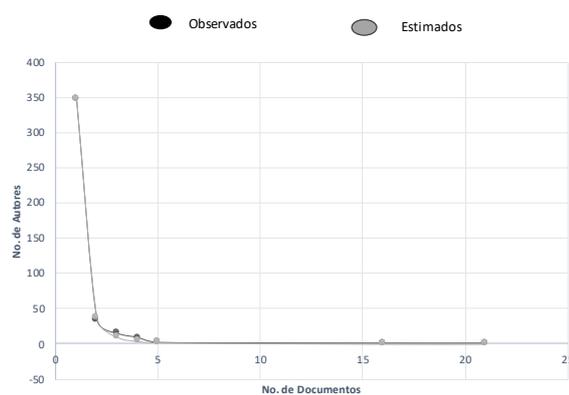


Gráfico 5. Productividad estimada vs observada

La productividad estimada se mantuvo en un rango bastante cercano con la productividad observada (Gráfico 5) mientras que los 216 documentos publicados sobre ontologías terminológicas fueron producidos por 420 autores.

En cuanto al número de artículos que los expertos publicaron, la relación número de artículos/autores se distribuyó del modo siguiente (Tabla IV):

Artículos	Autores esperados	Autores observados
1	354	354
2	42	38
3	12	15
4	5	9
5	2	2
16	0	1
21	0	1

Tabla IV. Número de artículos por autor

Según los datos, 354 autores (84 %) publicaron solo un trabajo sobre ontologías terminológicas; 64 autores (15,5 %) registran entre 2 y 5 estudios y sólo 2 autores (0,5 %) cuentan con más de cinco investigaciones sobre el tema. Tales resultados coinciden con la teoría de Ramsden (1994) quien afirma que la producción científica de los investigadores en una línea de trabajo suele ser baja, debido a la naturaleza de la investigación científica, los intereses mutables de estudio y las exigencias de tratamiento de temas o líneas de investigación a los que, frecuentemente, los académicos están sujetos en sus instituciones de afiliación.

Los autores con mayor producción científica sobre ontologías terminológicas son (Gráfico 6):

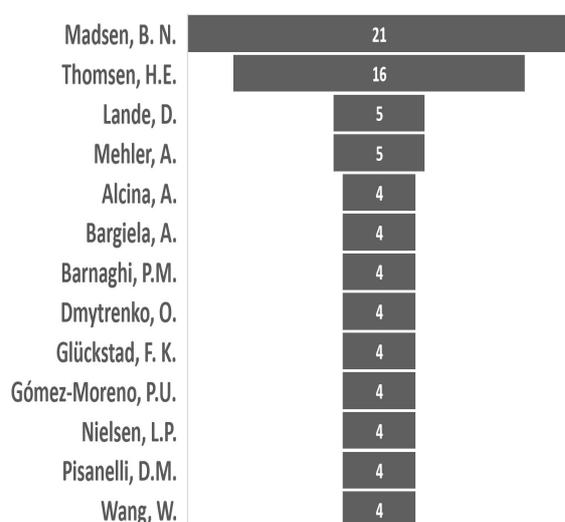


Gráfico 6. Número de trabajos por autor

La autora más productiva es Bodil Nistrup Madsen, reconocida investigadora de la Escuela de Negocios de Copenhagen. Sus líneas de estudio engloban la terminología, los sistemas de gestión terminológica, las bases de datos, la terminología para el modelado de datos, las ontologías y la representación y organización del conocimiento. En vida, Bodil Nistrup Madsen (1948-2022) recibió el reconocimiento de profesora emérita en su centro de trabajo y obtuvo dos renombrados premios internacionales: en 2002 recibió el Premio de la Fundación Hedorf a la Investigación en Lenguaje para Aplicaciones Específicas, y en 2006 recibió el premio Eugen Wüster en reconocimiento a sus investigaciones y proyectos sobre terminología aplicada.

La segunda autora con mayor productividad sobre ontologías terminológicas es Hanne Erdman Thomsen. Sus líneas de investigación son ciencias computacionales y lingüística aplicada. Ha laborado en la Universidad de Copenhagen como asistente de investigación y en la Escuela de Negocios de Copenhagen en el departamento de lingüística computacional. Actualmente se encuentra adscrita al proyecto Danish Taxation en tareas de arquitectura de la información.

Si consideramos la teoría de producción científica establecida por Braun, Glänzel y Schubert (2001) y adaptada por Walter, Lanza, Sato, Silva y Bach (2010) la relación autores/número de trabajos puede ser agrupada en categorías:

- *Los de única vez*: autores que en un periodo temporal establecido sólo han publicado un documento.

- *Retirantes*: autores con al menos dos documentos publicados, pero ninguno en los últimos tres años.
- *Entrantes*: autores con al menos dos documentos publicados en últimos tres años del periodo establecido.
- *Transeúntes*: autores con al menos dos documentos publicados, pero al menos un documento publicado en los últimos tres años del periodo de estudio.
- *Continuantes*: autores con al menos cinco documentos publicados y al menos uno de ellos publicado en los últimos tres años del lapso de estudio.

Con base en tal categorización, los autores que publican sobre ontologías terminológicas responden a la siguiente tipología (Tabla V):

Categoría	Autores	Porcentaje
Los de única vez	354	84 %
Retirantes	47	11 %
Entrantes	11	3 %
Transeúntes	6	1,5 %
Continuantes	2	0,5 %

Tabla V. Categorización de autores

El 84 % de los autores sólo ha publicado un documento sobre el tema, mientras que hay un 11 % de autores retirantes y un 3 % de entrantes. Tal hecho implica que el tema carece de un número considerable de autores cautivos, cuya línea de investigación se centre en ontologías terminológicas.

3.7. Colaboración

En la producción científica sobre ontologías terminológicas también se analizó la colaboración entre autores, según el coeficiente de colaboración propuesto por Ajiferuke, Burell y Tague (1988). Los datos evidenciaron trabajos producidos por 1 y hasta 13 autores (Tabla 6, en la página siguiente).

La colaboración entre académicos de una misma disciplina o de disciplinas diversas posibilita alcanzar un fin común y en los últimos años se ha convertido en una práctica que eficiente los recursos asignados a los investigadores y las instituciones. En adhesión, la colaboración es una práctica que enriquece a las comunidades científicas a través de la comunicación y el enriquecimiento entre pares (González Alcaide y Gómez Ferri, 2014, p. 4):

Ciencia y colaboración se han constituido en un binomio indisoluble para posibilitar el progreso y el avance del conocimiento [...] la colaboración constituye uno de los rasgos definitorios de la evolución de la Ciencia a lo largo de las últimas décadas.

Quinquenio	Número de publicaciones con 1, 2, 3, n autores											CC
	1	2	3	4	5	6	7	8	9	12	13	
1997-2001	0	1	2	2	0	0	0	0	0	0	0	0.67
2002-2006	7	5	10	3	0	1	0	0	0	0	0	0.47
2007-2011	24	18	16	4	4	0	0	0	1	0	0	0.40
2012-2016	17	20	9	6	5	2	1	0	0	1	1	0.47
2017-2022*	14	15	11	10	4	2	0	0	0	0	0	0.49
Total	62	59	48	25	13	5	1	0	1	1	1	0.46

* Dado que la investigación abarca hasta 2022, el último grupo se formó de seis años.

Tabla 6. Colaboración entre autores

En el corpus analizado, se detectó la siguiente tendencia de coeficiente de colaboración por quinquenios (Gráfico 8).

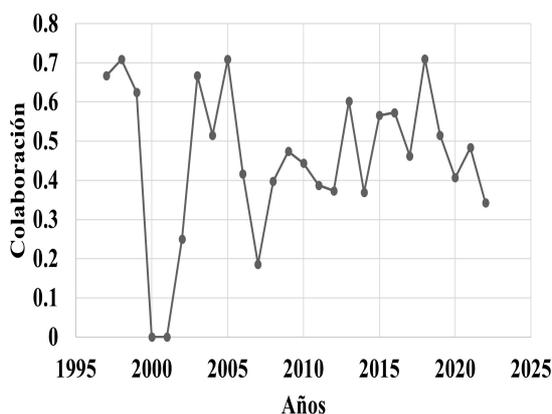


Gráfico 7. Colaboración entre autores por años

El estado de la colaboración en la literatura sobre ontologías terminológicas ha pasado por varios momentos. Al principio (1997-2005) el índice de colaboración se mantuvo con tendencia alta; sin embargo, en los últimos años (2005-2021) ha decrecido ligeramente.

Tal hecho se debe a que, cuando las ontologías se introdujeron en la bibliotecología y otras ciencias de la información, se les observó como sistemas altamente complejos que para su estudio e implementación requerían trabajo multidisciplinario y conocimiento experto de lingüistas, documentalistas e informáticos. Luego, en las últimas décadas el perfil de los bibliotecólogos y otros profesionales de la información se ha enriquecido. Hoy en día tales expertos cuentan con mayor formación sobre terminología, especialmente enfocada a la construcción de lenguajes documentales y experiencia en el manejo de bases de

datos relacionales y programas para la codificación de sistemas de organización del conocimiento, lo que reduce el número de autores en proyectos vinculados con ontologías terminológicas.

Pese a lo señalado anteriormente, la colaboración puede asumirse como una característica esencial de la producción científica sobre ontologías terminológicas. Al respecto es importante retomar a Almeida y Costa (2021) quienes señalan que los retos que plantea el desarrollo de los recursos terminológicos para la web semántica han propiciado el estrechamiento de vínculos entre la terminología, la ingeniería y las ciencias documentales, en una búsqueda por insertar los vocabularios de diversos campos, desde las ciencias médicas hasta el patrimonio cultural, en el marco de la web semántica y una recuperación más eficaz en la búsqueda temática o por palabras clave.

4. Conclusiones

Las tecnologías de la web semántica han ganado gran atención en los últimos años y para apoyarlas han surgido recientes herramientas y técnicas. La utilización de ontologías es una de las principales características de las bibliotecas digitales semánticas y, en particular, las ontologías terminológicas son un soporte fundamental como instrumentos para definir la terminología que represente temáticamente recursos y campos de conocimiento mediante conceptos, clases, atributos y relaciones.

Considerando lo anterior, la producción científica sobre organización de la información y el conocimiento desde principios del siglo veinte ha registrado una creciente producción científica sobre ontologías terminológicas. De esos años a la fecha la literatura sobre el tema se ha mantenido como un tema de interés: las ontologías han mostrado ser una herramienta útil para la representación y organización terminológica-temática de dominios de conocimiento, eliminando los problemas de ambigüedad, proporcionando interrelación y semántica a la terminología del área.

En términos de investigación, se han revisado los fundamentos teóricos de las ontologías terminológicas; se han construido modelos terminológicos que cubren diversas áreas de conocimiento (medicina, humanidades, arte, etc.) y muchas ontologías han sido integradas en sistemas de recuperación. Pese a ello, aún persiste un cuestionamiento esencial sobre: ¿cuál será el futuro de las ontologías terminológicas en funciones de organización del conocimiento?

A partir de la investigación realizada es posible concluir que las ontologías son un tema de estudio con tendencia de crecimiento estable. Actualmente, el tópico mantiene el interés de un grupo de autores realmente comprometidos con el estudio del tema, quienes han descubierto que el uso de modelos axiomáticos proporciona una semántica adicional y una explicitación formal de relaciones de gran utilidad en la representación, organización y recuperación de la información.

En la producción literaria sobre el tema, predomina la publicación en lengua inglesa y las actas de conferencia como recurso dominante. Basado en ello, las tendencias futuras nos permiten proponer una ampliación de las ontologías terminológicas en contextos locales, en publicaciones en lengua española, revistas locales y encuentros regionales donde los expertos, que asisten a convenciones internacionales, puedan tener contacto no sólo con investigadores sino con bibliotecarios, que día a día laboran en unidades de información en las cuales las ontologías terminológicas puedan tener implementaciones propositivas.

Un segundo rasgo de la literatura sobre ontologías terminológicas es la colaboración, dicha característica se observa como elemento fundamental para ampliar al panorama de aplicación de las ontologías terminológicas en funciones de indexación de recursos, organización de dominios de conocimiento y recuperación de información. Los resultados del artículo coinciden con el trabajo previo de Zhu, Kong, Hong, Li y He (2015) en el que se argumentó sobre la colaboración y el trabajo interdisciplinario como aspecto clave en la investigación sobre ontologías.

Los resultados obtenidos en la presente investigación coinciden con lo señalado por García Marco (2008) cuando, hacia 2007, apuntaba una etapa de despegue de las ontologías. En adhesión también es posible mencionar que, de esos años a la actualidad, la producción científica sobre ontologías ha entrado en una etapa de estabilización que ya no registra las cifras efervescentes de un primer interés sobre el tema.

Finalmente, es posible afirmar que las ontologías han trastocado el mundo de la organización del conocimiento y son un tema interdisciplinar de gran interés para documentalistas, lingüistas, filósofos y expertos en ciencias computacionales. Como señalan Netto y Lima (2017) las ontologías aún tienen muchas aristas de estudio pendientes. Por tanto, se espera que su tendencia de investigación se mantenga constante tanto en la determinación de sus aspectos teóricos como en la definición de sus implementaciones en sistemas de información.

Notas

- (1) En Scopus y Web of Science fue posible llevar a cabo la búsqueda de los descriptores en los campos de título, temas y resumen. En Redalyc, Dialnet y WordCat fue posible buscar en los campos de título y palabras clave. Mientras que en Google Académico sólo se llevó a cabo la búsqueda en palabras clave.
- (2) La estimación del periodo de duplicación de la literatura es precisa en términos estadísticos; sin embargo, se observa algo incierta, debido a que los años 2021 y 2022 registraron baja producción sobre el tema (5 y 6 documentos, respectivamente) e incidieron, de manera negativa, en la tendencia de crecimiento que registraba 12 documentos anuales en promedio. Tal hecho, unido a que esos años, se caracterizaron por una pandemia mundial derivada del COVID-19 y la suspensión de encuentros de la especialidad, que son la mayor fuente de producción literaria, son un aspecto negativo que, probablemente, no esté ofreciendo datos completamente confiables. De este modo, se asume que: es probable que el tiempo de duplicación de la muestra sea algo menor al estimado en el artículo.

Referencias

- Ajiferuke, Isola; Burell, Q.; Tague, Jean (1988). Collaborative coefficient: a single measure of the degree of collaboration in research. // *Scientometrics*. 14:5/6, 421-433.
- Almeida, Bruno; Costa, Rute (2021). OntoAndalus: an ontology of Islamic artefacts for terminological purposes. // *Semantic Web*. 12:2, 295-311.
- Almeida, Mauricio; Bax, Marcello (2003). Uma visão geral sobre ontologias: pesquisa sobre definições, tipos, aplicações, métodos de avaliação e de construção. // *Ciência da informação* 32, 7-20.
- Aparicio, Gloria; Txomin Iturralde; Maseda, Amaia (2019). Conceptual structure and perspectives on entrepreneurship education research: a bibliometric review. // *European research on management and business economics*. 25:3,105-113.
- Braun, Tibor; Glänzel, Wolfgang; Schubert, Andrés (2001). Publication and cooperation patterns of the authors of neuroscience journals. // *Scientometrics*. 51:3, 499-510.
- Drubin, David; Kellogg, Douglas (2012). English as the universal language of science: opportunities and challenges. // *Molecular biology of the cell*. 23:8, 1399.
- Egghe, Leo; Ravichandra Rao, I.K (1992). Classification of growth models based on growth rates and its applications. // *Scientometrics*. 25:1, 5-46.
- García Marco, Francisco Javier (2008). El nacimiento y el despegue de la investigación moderna sobre ontologías. // *Anuario ThinkEPI*. 2:122-125.
- González Alcaide, Gregorio; Gómez Ferri, Javier (2014). La colaboración científica: principales líneas de investigación y retos de futuro. // *Revista Española de Documentación Científica*. 37:4, 1-15.
- Gruber, Thomas (1995). Toward principles for the design of ontologies used for knowledge sharing? // *International Journal of Human and Computer Studies*. 43:5-6, 907-928.
- Guarino, Nicola (1995). Formal ontology, conceptual analysis and knowledge representation. // *International Journal of Human-Computer Studies*. 43:5-6, 625-640.
- Hodge, Gail (2000). Systems of knowledge organization for digital libraries: beyond traditional authority files. Washington: The Digital Library Federation Council on Library and Information Resources.

- Lacasta, Javier; Nogueras-Iso, Javier; Zarazaga-Soria, Francisco Javier (2010). Terminological ontologies: design, management and practical applications. Nueva York: Springer.
- López Alonso, Miguel Ángel. (2000). Las estructuras conceptuales de representación del conocimiento en Internet. // Scire: representación y organización del conocimiento. 6:1, 107-123.
- Maletta, Héctor (2009). Metodología y técnica de la producción científica. Lima: CEPES-Universidad del Pacífico.
- Netto, Cristiane; Lima, Gercina (2017). Visualização de ontologias: estudos e perspectivas. // Informação & Sociedade. 27:3, 59-72.
- Pedraza, Rafael; Codina Lluís; Rovira Cristòfol (2007). Web semántica y ontologías en el procesamiento de la información documental. // El profesional de la información. 16:6, 569-78.
- Piedra Salomón, Yelina; Martínez Rodríguez, Ailín (2007). Producción científica. // INFO: ciencias de la información. 38:3, 33-38.
- Pisanelli, Domenico; Gangemi, Aldo; Steve, Geri (1997). WWW-available conceptual integration of medical terminologies: the ONIONS experience. // Masys, Daniel (ed.) Proceedings of the AMIA Annual Fall Symposium. Nashville, Tennessee: American Medical Informatics Association. 575-579.
- Ramsden, Paul (1994). Describing and explaining research productivity. // Higher education. 28:2, 207-226.
- Reitz, Joan (ed.) (2004). Dictionary for library and information science. Westport: Libraries Unlimited.
- Sánchez-Cuadrado, Sonia; Morato-Lara, Jorge; Palacios-Madrid, Vicente; Llorens-Morillo, Juan; Moreira-González, José-Antonio (2007). De repente, ¿todos hablamos de ontologías? // El profesional de la información. 16:6, 562-568.
- Suárez Sánchez, Adriana (2022). Ontologías: principios y aplicaciones en la bibliotecología. Ciudad de México: Instituto de Investigaciones Bibliotecológicas y de la Información, Universidad Nacional Autónoma de México.
- Spinak, Ernesto (1996). Diccionario enciclopédico de bibliometría, cienciometría e informetría. Caracas, Venezuela: UNESCO.
- Urbizagástegui Alvarado, Rubén (1999). La ley de Lotka y la literatura de bibliometría. // Investigación Bibliotecológica. 13:27,125-141.
- Urbizagástegui Alvarado, Rubén (2021). Análisis bibliométrico de la literatura sobre fósiles en el Perú, 1840-2020. // Revista del Instituto de Investigación de la Facultad de minas, metalurgia y ciencias geográficas. 24:48, 153-165.
- Van Heijst, G.; Schreiber, A.; Wielinga, B. (1997). Using explicit ontologies in KBS development. // International Journal of Human-Computer Studies. 45, 183-292.
- Walter, Silvana Anita; Lanza, Beatriz Barreto Brasileiro; Sato, Kawana Harue, Silva, Eduardo Damião; Bach, Tatiana Marceda (2010). Análise da produção científica de 1997 a 2009 na área de estratégia: produção e continuidade de atores e cooperação entre instituições brasileiras e estrangeiras. // Anais do Encontro da Associação Nacional dos Programas de Pós-graduação em Administração.34. São Paulo: ANPAD.
- Zhu, Qiaoli; Kong, Xuesong; Hong, Song; Li, Junli; He, Zongyi (2015). Global ontology research progress: a bibliometric analysis. // Aslib Journal of Information Management. 67:1, 27-54.

Enviado: 2023-03-02. Segunda versión: 2023-06-30.
Aceptado: 2023-11-02.

Fundamentos linguísticos da indexação: uma revisão

Fundamentos lingüísticos de la indización: una revisión

Linguistic fundamentals of indexing: a review

Alexandre Robson MARTINES, Carlos Cândido ALMEIDA

UNESP – Marília/SP, Brasil, alexandre.martines@unesp.br, carlos.c.candido@unesp.br

Resumen

Las etapas del proceso de indización exigen del profesional conocimientos técnicos, específicos y especializados de la propia práctica, así como del universo científico o cultural inherente al material a clasificar e indizar. Los aspectos terminológicos son insumos fundamentales para el desarrollo del lenguaje documental; el análisis del lenguaje se ha convertido en un enfoque necesario, dado que el estudio del lenguaje a través de la lingüística y la filosofía del lenguaje ha evolucionado considerablemente en los últimos años. El objetivo de esta investigación es discutir el proceso de indización y analizar los fundamentos de la lingüística y la filosofía del lenguaje que pueden contribuir a una comprensión epistemológica de las capas de conocimiento que rodean todo el proceso de indización. Se aplicó una metodología cualitativa, exploratoria y bibliográfica. La tradición inherente a la indización ha hecho posible el uso de pronombres interrogativos, aspectos notionales y facetas, entre otros muchos modelos. Sin embargo, los aspectos modales impulsan el formato de la indización, mientras que la estrategia que subyace al formateo de los modales en el tratamiento de la información es el modelo que repercute en el comportamiento del usuario a la hora de realizar búsquedas.

Palabras clave: Indización. Lingüística. Fundamentos lingüísticos. Modales.

1. Introdução

O processo informacional é muito complexo e envolve vários níveis de expressão, interpretação e comunicação. Muitas das práticas de organização e representação da informação remontam à antiguidade, quando as pessoas responsáveis pela sistematização da produção do conhecimento se utilizavam de etiquetas, placas de argila, ou qualquer outro tipo de evidência para indicar o conteúdo presente em um determinado conjunto de materiais. Graças a essas práticas peculiares, que no passado distante eram tão simples e na atualidade exigem uma reflexão profunda e contínua de sua aplicabilidade, que a humanidade pôde prosperar e aprender com as práticas e epistemes dos antepassados, como ainda estudar sobre eles.

Abstract

The stages of the indexing process demand from the professional technical, specific and specialized knowledge about the practice itself, in addition to the scientific or cultural universe inherent to the material to be classified and indexed. Terminological aspects are fundamental inputs for the elaboration of documentary language; the analysis of language becomes a necessary approach, since the study of language through linguistics and philosophy of language has shown a lot of evolution in recent years. The aim of this research is to discuss the indexing process and analyze fundamentals of linguistics and the philosophy of language that can contribute to the epistemological-logical understanding of the layers of knowledge that involve the entire indexing path. A qualitative, exploratory and bibliographic methodology was applied. The tradition inherent to indexing made possible, among so many models, the interrogative pronouns, the notional aspects, facets, however the modal aspects lead the indexing format, since the strategy behind the formatting of the modals in the treatment of information is the model that impacts on the user behavior when carrying out searches.

Keywords: Indexing. Linguistics. Modals. Linguistic fundamentals.

Nesse cenário, todas as etapas que envolvem o processo de indexação exigem do profissional conhecimento técnico, específico e especializado sobre a própria prática, como ainda sobre o universo científico ou cultural inerente ao material a ser classificado e indexado. Frente a isso, observa-se que além da informação manifestada no enunciado dos artefatos, estes, durante o tratamento, são analisados como textos, sejam verbais, sejam não verbais, também são explorados a linguagem, os recursos tecnológicos e fatores psico-sociocultural do autor do texto, das formas de vida que envolvem a discussão proposta pelo texto, do profissional que realiza a indexação e do usuário.

Ademais, aspectos terminológicos são insumos fundamentais para a elaboração de linguagem

documental, como por exemplo os tesouros, assim tornando-se elementos constituintes do processo e do produto decorrentes da prática de indexação. Nessa linha, a análise da linguagem torna-se uma abordagem necessária, visto que o estudo da linguagem através da linguística e da filosofia da linguagem tem apresentado bastante evolução nos últimos anos.

Pode-se entender que a *filosofia da linguagem* remonta à antiguidade grega, já discutida com Platão e Aristóteles, retomada por Santo Agostinho, explorada por segmentos mentalistas como a escola de Port-Royal; por segmentos racionalistas, como os defendido por R. Descartes, G. W. Leibniz; por segmentos empiristas, como os apresentados por F. Bacon, T. Hobbes, G. Berkeley, J. Locke, D. Hume; além de ser alcançado pelos pensamentos iluministas como E. B. Condillac, D. Diderot, pelo lado francês, enquanto I. Kant, C. Wolff, J. H. Lambert, pela corrente alemã. A filosofia da linguagem também encontra lugar entre os lógicos e os estéticos do século XIX, como por exemplo J. G. Fichte, G. W. F. Hegel, W. von Humboldt, B. Bolzano, L. V. Welby, C. K. Ogden, I. A. Richards, F. Schiller, F. W. J. Schelling, C. S. Peirce. No século XX, o tema chamou a atenção de filósofos e de lógicos, como G. Frege, R. Carnap, J. L. Austin, J. R. Searle, L. Wittgenstein, J. Habermas e muitos outros estudiosos de MIT (Massachusetts Institut Technology), alguns contribuindo para avanços da linguística, como N. Chomsky e J. Fodor.

O problema da pesquisa considera que, embora a filosofia da linguagem tenha apresentado muitas contribuições para o estudo da linguagem, aspectos que envolvem sua forma, sua função como aparato da representação da realidade, da mediação da realidade, da descrição da realidade, da cognição, da relação mente, forma e realidade, acesso ao conteúdo e suas múltiplas plataformas de significação. De modo geral, o estudo da linguagem no cenário da ciência da informação, mais especificadamente do campo da *organização do conhecimento* no Brasil e na Espanha tem se fixado em uma terminologia ancorada na fundamentação apresentada no positivismo lógico, decorrente do Círculo de Viena e na semiologia de F. Saussure, de 1916, a qual se apega à concepção do valor semântico oriundo da *langue*.

Frente a isso, o objetivo desta pesquisa é discutir processo de indexação e analisar fundamentos da linguística e da filosofia da linguagem que possam contribuir para a compreensão epistemológica das camadas de saberes que envolvem todo o percurso de indexação.

Para tanto, aplicou-se uma metodologia de natureza qualitativa, bibliográfica, já que busca analisar o estado da arte nas diversas correntes teóricas de concepções sobre a linguagem com o fito de evidenciar uma base teórica de nível internacional para validar sua aplicabilidade na relação mente, realidade e representação. Assim, foram consultadas as bases de dados: Brapci, Dialnet, Scopus, Google Scholar, além de anais de eventos como Enancib, Ibersid, ISKO Brasil, ISKO ibérico. A coleta se deu através de palavras-chave como *indexação*, *representação da realidade*, *representação da informação*, *processo cognitivo*, *conceito*, *terminologia*, *linguagem documental*, *linguagem de indexação*, *práticas de indexação* e *tesouros*. Recuperaram-se materiais em português, espanhol, inglês e francês. Diante do material consultado, destacam-se: García Gutiérrez (1985); Izquierdo Arroyo (1990, 1993, 1995); Gil-Leiva & Rodríguez Muñoz (1996); Fujita & Gil-Leiva (2009, 2014); assim como Córrea & Lapa (2013); Fujita (2016, 2021); Gil-Leiva (2021), para tratar dos processos, das práticas e da política de indexação; Nöth (1996); Saussure (2006); Zilberberg (2006); Fiorin (2007); Mussalim & Bentez (2011, 2012); Milani (2016); Nöth & Santaella (2017) para tratar de linguagem, linguística e texto.

Com o resultado desta pesquisa, espera-se que haja uma ampliação na discussão sobre o papel da linguagem no processo de representação da informação, em destaque no processo e prática de indexação, visto que a filosofia da linguagem e as teorias da linguagem evoluíram suas concepções epistemológicas e já se entende que a linguagem não é apenas a forma, mas atua na constituição do conteúdo, o que só é possível porque a linguagem é um dos fundamentos da cognição.

2. Indexação: práticas, modelos e política

A indexação é um processo complexo no cenário da ciência da informação e, conseqüentemente, para o campo da organização do conhecimento, visto que sua atuação engloba fatores herdados da ciência da documentação e da biblioteconomia, tendo em vista as necessidades informacionais oriundas da prática de classificar e catalogar, bem como o processo de elaboração de instrumentos descritivos cuja função é garantir a recuperação de informações de modo eficiente, rápido, preciso e competitivo.

A indexação como processo está fundamentada em estágios preciosos para que a evolução do serviço acompanhe as demandas sociais e científicas. Frente a uma renovação tecnológica, os

aparatos que circundam a prática de indexação também estão em constante debate visando à renovação.

A evolução da prática de indexação demonstra que as preocupações foram se tornando mais complexas, principalmente diante da quantidade de informações produzidas na atual sociedade, além de a indexação assumir nessa nova atuação um importante papel de democratização da informação, isto é, se no passado a preocupação estava centrada nos profissionais, ou melhor, a indexação era um mecanismo de navegação utilizada pelos profissionais para acessar os sistemas informacionais, portanto uma prática voltada para sistemas fechados, como as bibliotecas; agora, com o advento da tecnologia, a indexação enfrenta as exigências de um mundo conectado e em constante comunicação.

Nessa linha, entende-se que a linguagem de indexação, como ainda os comandos de navegação, ou os motores de busca têm se adequando às inovações dos sistemas de informação, ou seja, o processo de informatização e o uso de algoritmos têm possibilitado novos comandos e, conseqüentemente, novos modelos de conduta não só do profissional, mas também do usuário, já que este, em diversos sistemas informacionais, possui autonomia para realizar a sua busca.

Sendo assim, o modelo de indexação voltado para as bibliotecas necessita incluir novas propostas de reflexão a fim de avaliar se o modelo de indexação continua condizente com a conduta de quem realiza a recuperação da informação, pois é preciso ter em conta que o usuário se tornou plural, isto é, não se trata apenas do profissional da informação, do cientista ou do especialista, ou ainda do estudante. Com as novas propostas de acesso à informação, qualquer pessoa pode acessar a sistemas informacionais, assim a recuperação da informação não estará concentrada na terminologia especializada, pois há determinados grupos de usuários que não possuem conhecimento específico e a busca pode ter caráter introdutório no assunto.

Frente a esses fatores técnico-sociais, é objetivo desta seção retomar alguns fundamentos que envolvem o processo e a prática de indexação, a fim de analisar seus procedimentos, etapas e política para refletir sobre a relação indexação e usuário.

Dessa forma, entende-se que a informação permeia todo o processo e sua natureza é constituída em camadas de significado e significação, já que há versões sobre a mesma informação, ou seja, há a informação em seu estágio produtivo-discursivo, que seria sua condição de informação

produzida no interior de um documento, sob a intencionalidade de seu produtor/autor/escritor.

Há um segundo estágio de investigação, em que um profissional da informação, responsável pela sua organização e representação interage com o documento, assim havendo o momento de compreensão do assunto, do recorte temático, da intencionalidade, da aspectualização, para inserir esse documento em um contexto, ou seja, a chamada focalização, em que se aponta os conceitos principais, aqueles que melhor traduzirão o conteúdo do documento para potenciais usuários.

Um terceiro estágio, a condensação, em que todo o discurso do texto é canalizado na combinação de algumas palavras, no caso termos especializados, ou através de frases, conjuntos que são conduzidos pela linguagem documental; por fim, o estágio da comunicação, em que o usuário, através de comandos linguísticos, estabelece uma congruência semântico-linguística entre a interface utilizada pelo usuário e o conjunto de recursos indexados no sistema informacional.

Devido a esses fatores, a prática de indexação não se resume simplesmente a uma atividade de leitura, interpretação e seleção de palavras, embora essas palavras sejam correlatas a uma terminologia especializada.

As estratégias que envolvem a indexação precisam funcionar como âncoras de comunicação entre sistema informacional e usuário, por isso é preciso haver um modelo de indexação, um roteiro que possibilite que as informações estratégicas sejam veiculadas, pois o papel da indexação não é a tradução do texto original em um objeto informacional, já que isso é impossível, tendo em vista a disposição discursiva do documento original, assim é necessário entender que esse roteiro explora fundamentos modais da língua, os quais articulam informações no nível cognitivo de qualquer pessoa que interaja com ele.

Consoante Gil Leiva (2012, p. 31, tradução nossa),

a indexação gera palavras-chave, índices ou os cabeçalhos de assunto de um documento. Para obtê-los, previamente foi desencadeada uma sucessão interativa e simultânea dos processos mentais que têm a ver com a percepção da informação, da memória e da compreensão.

Frente a isso, reconhece-se que a aplicação das palavras-chave para identificar o documento é o produto da indexação, o qual é decorrente de etapas em que o indexador assume o papel de leitor, o qual age diante de um processo cognitivo que envolve conhecimento da técnica e do conteúdo, a linguagem, a percepção, a memória, a tomada de decisões e o autocontrole frente suas

escolhas visando às estratégias de recuperação da informação que estão no devir.

De todo modo, a prática de indexação se perfaz em três momentos: a) leitura e interpretação dos documentos; b) o reconhecimento e estabelecimento do assunto, dos temas e da intencionalidade presentes no texto, desse modo compreendendo as linhas de conhecimento aplicadas e, assim, identificar e selecionar os conceitos; c) a representação dos conceitos selecionados fazendo uso de uma linguagem documental, ou seja, aplicando termos fixados em uma linguagem determinada de indexação ou de um tesouro (Unesco, 1981).

Ao utilizar um tesouro como instrumento de indexação, a vantagem é a condução do processo por vocabulário controlado, bem como a correlação semântica ser orientada pelas relações dos termos. Com isso (Unesco, 1981, p. 91, tradução nossa),

[...] o número de termos indexadores de um documento e o desdobramento de entradas podem ser reduzidos sem danos, pois as relações genéricas e outras podem ser estabelecidas diretamente do próprio tesouro”.

Conforme aponta Fujita (2016, p. 4), há “[...] mudanças significativas no comportamento informacional de acesso e uso do catálogo por usuários que solicitam recuperação por assunto com mais especificidade, compatibilidade com sua linguagem de busca e disponibilidade de mecanismos de interação”, fator que demonstra uma perspectiva de indexação para o profissional que indexa, como uma outra potencial para o usuário, decorrente do estágio cognitivo que envolve conhecimento, motivações contextuais no momento da indexação ou da busca, como ainda a escolha da linguagem.

Nessa perspectiva, Fujita (2016, p. 4), indica que “[...] a linguagem documental é o instrumento que faz a diferença na mediação com a linguagem do usuário durante as estratégias de buscas em catálogos online”. Além disso, diante da amplitude de pesquisas realizadas através da tecnologia e bancos de dados online, afirma-se que “a qualidade dos serviços de disseminação online da biblioteca depende da linguagem utilizada para a indexação e recuperação porque não é possível a mediação física do bibliotecário, sobretudo quando a busca é por assuntos e exige mais precisão ou revocação com alta especificidade ou exaustividade” (Fujita, 2016, p. 4).

Diante disso, a indexação precisa lidar com o ponto mais frágil de sua atuação frente à organização e à representação da informação, visto que diante da pluralidade de meios de acesso e a democratização do perfil do usuário, os modais

de indexação possivelmente não sejam satisfatórios para contemplar os métodos de busca de modo amplo e geral, pois “para que essa mediação aconteça é necessário que o bibliotecário realize o processo de indexação na catalogação com uso de uma linguagem documental que seja a mesma a ser utilizada pelo usuário no momento da busca” (Fujita, 2016, p. 4) para, assim, garantir que haja o mesmo princípio linguístico que sustente a representação da informação e a correlação ao conteúdo registrado no sistema informacional, isto é, um padrão cada vez mais difícil de se manter.

Assim, a fim de evitar problemas na recuperação da informação, o processo de indexação é visto como elemento essencial de uma política de tratamento temático da informação, sendo influenciada por algumas variáveis como a escolha da linguagem, exaustividade, especificidade, revocação e precisão.

Para Fujita e Gil-Leiva (2014, p. 51), a linguagem é ferramenta de conversão, isto é, trata-se do “instrumento da indexação que cumpre a função de controle de vocabulário para o indexador e deve realizar a mediação na recuperação por assuntos pelo usuário”.

Assim, a exaustividade é componente da política de indexação e “[...] corresponde à quantidade de termos atribuídos à representação da forma que sejam suficientes para o âmbito de abrangência do conteúdo documental” (Fujita & Gil-Leiva, 2014, p. 51).

Por sua vez, a especificidade também é considerada componente da política de indexação e é “[...] determinada por níveis de especificidade entre os termos atribuídos para representar a abrangência do conteúdo para que determinado termo atribuído possa ser mais específico ou incluir outro mais específico” (Fujita & Gil-Leiva, 2014, p. 51).

Por fim, a revocação e a precisão são fatores dependentes e entre elas “[...] existe uma relação de relevância necessária para a recuperação da informação a fim de satisfazer uma necessidade de informação”, visto que um é responsável pela recuperação de informações, por conseguinte documentos úteis, enquanto a outra tem a função de evitar os inúteis (Fujita & Gil-Leiva, 2014, p. 51).

Nessa correlação de instrumentalização, evidencia-se que algumas práticas são conduzidas por uma normalização, assegurada por uma política de atuação; outros fatores são conduzidos pela matemática, dessa forma a fundamentação da lógica contribui para estabelecer parâmetros. No entanto, o maior problema desse processo é a

linguagem, visto que quanto mais abrangente for o sistema informacional em relação à indexação de informações armazenadas e maior for a variedade do perfil democrático do usuário, menor será a compatibilidade linguística. Desse modo, embora os tesouros tenham uma tensão semântica bastante forte, não serão suficientes para impedir esse distanciamento, devido à diversidade entre buscas conduzidas por uso de terminologias especializadas e buscas conduzidas por linguagem natural.

Sendo assim, a política de indexação deve conter esses fatores, já que o processo de indexação exige três fundamentos basilares: a) a política: sistematizada por normas e convenções; b) o sistema de indexação e recuperação: organizado e executado por lógica; e c) a linguagem: o insumo que alimenta o sistema informacional e também a interface que conecta documento, informação e usuário.

Como aponta Fujita (2021, p. 17), “a política de indexação não deve ser vista como uma lista de procedimentos a serem seguidos, e sim um conjunto de decisões que esclareçam os interesses e objetivos de um sistema de informação”. Outrossim, “a política decide não só sobre a consistência dos procedimentos de indexação em relação aos efeitos que se necessita obter na recuperação, mas, principalmente, sobre a delimitação de cobertura temática em níveis qualitativos e quantitativos tendo em vista os domínios de assuntos e as demandas” (Fujita, 2021, p. 17).

Somado a isso, Fujita e Gil-Leiva (2009, p. 156) apontam que a política de indexação deve ser entendida como [...] uma filosofia que reflita os interesses e objetivos da unidade de informação”, o que impacta nas questões gerenciais e estratégicas inerentes à configuração do contexto em que se apresentam as unidades de informação. [...] Com isso,] a indexação e, por conseguinte sua política, deve fazer parte do planejamento dos sistemas de informação de modo a compor-se como um conjunto de procedimentos, materiais, normas e técnicas orientadas por decisões que refletem a prática e princípios teóricos da cultura organizacional.

Nessa perspectiva, entende-se que o processo de indexação apresenta um modelo de práticas, serviços e estratégias bastante estabilizados e consagrados em referência ao que conduz o trabalho do profissional da indexação atuando frente a sistemas informacionais cuja interação com os usuários é restrita ao conhecimento sobre os assuntos e temas pesquisados.

Todavia, decorrente aos novos desafios que o cenário da informação tem enfrentado com as novas tecnologias e novas práticas de acesso, novos bancos de dados e novos perfis de usuário, a indexação também é provocada a refletir

sobre suas práticas e como pode corroborar a democratização do acesso informacional de modo mais abrangente.

Fica evidente que a tradição envolta à indexação se perfaz por fundamentos criteriosos e sistematizados em normas, regras e políticas para que o procedimento seja realizado da melhor forma possível. Além de debates concernentes à ética, à correspondência informacional entre documento e objeto informacional, estudos e debates sobre a indexação também apresentam elementos da matemática e da lógica para nortear suas análises e avaliações referentes ao trabalho, ao processo, às demandas e aos sistemas.

Porém, fica evidente que o ponto mais frágil no complexo processo que envolve a indexação é a linguagem, visto que esta reflete fatores culturais, cognitivos, lexicológicos, semânticos, pragmáticos, semióticos e discursivos presentes em todos os estágios de tratamento, organização, representação e acesso da informação, e a aplicação de tesouros não tem se mostrado suficiente para garantir o controle dos documentos recuperados nas pesquisas realizadas pelos usuários.

3. Fundamentos linguísticos: tratamento da linguagem e interpretação

A indexação é processo e produto inserido em uma complexidade de fatores que envolvem a linguagem, o conhecimento e a informação. Nessa linha, a linguagem ganha destaque nessa discussão devido aos fatores linguísticos que são inerentes à terminologia, no entanto pouco se explora os diversos fundamentos da linguística na ciência da documentação, na biblioteconomia e no campo da organização do conhecimento, restringindo-se, na maior parte das abordagens, à semiologia saussureana e os padrões formalistas estabelecidos pelo positivismo lógico do Círculo de Viena, que nega a substância e a metafísica na composição da linguagem.

Sendo assim, é válido argumentar que os estudos da linguística avançaram consideravelmente referente a seu objeto. Inicialmente, era focado na *langue* saussureana, porém, com o passar dos anos, passou a explorar não somente a forma, mas caminhos para analisar a significação através da forma, bem como caminhos para acessar o conteúdo, para tanto se aproximando, de fato, da psicologia, da neurociência, da sociologia e da filosofia.

Pode-se destacar três grandes linhas de desenvolvimento da linguística no cenário mundial. A saber, a primeira linha, que se originou nas discussões oriundas dos círculos de Viena, de Praga, de Copenhague e de Paris. Inspirados

pela filosofia dos mentalistas, dos formalistas e do, até então recente, estruturalismo positivista, os estudiosos dessa linha propuseram, em destaque, a forma. Assim, surge a discussão sobre o valor da palavra e como esse valor é produzido e recebido na sociedade. Nessa linha, evidenciase a semiologia de F. Saussure, que no *Cours de Linguistique Générale*, em 1916, lança a discussão mais aprofundada sobre o signo linguístico, sua constituição como significante (imagem acústica) e significado (conceito), além das questões em torno da *langue* (língua como sistema) e da *parole* (fala como uso individual) (Nöth, 1996; Saussure, 2006; Fiorin, 2007; Mussalim & Bentez, 2011; Milani, 2016; Nöth & Santaella, 2017).

Os avanços de L. Hjelmslev, na década de 1930, para a linguística estrutural merecem destaque, principalmente na sua nova proposta de configuração dos aspectos semióticos da palavra, assim denominados em sua glossemática de plano da expressão e plano do conteúdo, além dos desdobramentos abordando a forma e a substância, ou seja, atribuindo os fundamentos de forma da expressão e substância da expressão para tratar a forma em seus insumos e seus produtos, bem como forma do conteúdo e substância do conteúdo direcionados ao tratamento do conteúdo na concepção nuclear da palavra e sua aplicação em uma sentença, assim ampliando a concepção de língua para os *functivos*, cuja junção promove uma semiótica e os fenômenos de sistema e processo (Nöth, 1996; Fiorin, 2007; Mussalim & Bentez, 2011; Nöth & Santaella, 2017).

Na sequência, os trabalhos de R. Barthes apresentam grandes contribuições para a semiologia, avaliando o impacto social dos signos e a configuração de valor linguístico, o qual é denominado de mito pelo linguista francês. Com isso, há uma terceira concepção sobre a análise dos objetos linguísticos, ou melhor, com Saussure a análise recaía na palavra, com Hjelmslev, houve um avanço para a sentença, já com o Barthes há a diferença entre o enunciado e o discurso. Evidentemente, os termos ainda não eram recorrentes nas discussões de Barthes, todavia com as reflexões acerca do papel do mito na sociedade e como este velava os fatos reais, Barthes já demonstrava que havia níveis de significação para além daquilo que era manifestante no nível do enunciado (Nöth, 1996; Fiorin, 2007; Mussalim & Bentez, 2011; Nöth & Santaella, 2017).

Nos anos de 1960, a linguística já se consolidara em outras linhas, com outras abordagens, entretanto na linha estruturalista, surgira A. J. Greimas, com sua semântica estrutural, em 1978. Estudos de sua fase estruturalista já abordavam a importância do nível discursivo e, inspirado na fi-

losofia clássica, melhor especificando, na filosofia lógica de Aristóteles, Greimas apresenta o percurso gerativo de sentido, ancorado em três níveis: profundo, narrativo e discursivo. Greimas deu continuidade a seu trabalho até os anos 1990 e durante esse período foi aperfeiçoando sua concepção de semiótica. Seus discípulos deram continuidade ao seu trabalho semiótico explorando, por exemplo, a semiótica plástica (Floch, 1985), semiótica tensiva (Zilberberg, 2015) e práticas semióticas e semiótica discursiva (Fontanille, 2012), entre tantos outros especialistas e trabalhos, essa linha da semiótica permanece produzindo teorias, métodos e práticas para tratar da forma, do conteúdo, do sentido, do significado, da significação, da relação linguagem e cognição (Nöth, 1996; Zilberberg, 2006; Fiorin, 2007; Mussalim & Bentez, 2011; Nöth & Santaella, 2017; Lopes & Hernandez, 2021).

Evidentemente que essa linha da semiologia e da semiótica influenciou outras tantas correntes da linguística, como por exemplo algumas abordagens da semântica, a linguística funcionalista, a linguística discursiva e textual, a sociolinguística, a lexicografia, a lexicologia, a morfologia, a terminologia, a linguística histórica, entre outros. No entanto, os estudos na ciência da informação direcionados às linguagens documentais continuam tendo como referência a semiologia de F. Saussure e os princípios formais do Círculo de Viena, a destacar os postulados de Frege, Russell e Carnap. É fato que a semiologia é precursora, mas também é fato que muitos avanços importantes sobre a significação já deveriam estar incorporados nos estudos do campo.

Na Espanha, há estudos motivados pelas teorias de L. Hjelmslev, como de García Gutiérrez (1984), Izquierdo Arroyo (1990, 1993, 1995), Izquierdo Alonso e Moreno Fernández (2010) e Izquierdo Alonso; Moreno Fernández e Sánchez Domínguez (2012); há também estudos que abordam a semântica estrutural ou traços da semiótica de Greimas, como Izquierdo Arroyo; outros estudos, como de Gil-Leiva que explora fundamentos da linguística textual e da linguística discursiva na vertente cognitiva de Van Dijk, Beaugrande e Dressler (Gil-Leiva, 2021).

Já no Brasil, os estudos são experimentais no que tange a reconhecer o alcance e limites dessas teorias linguísticas, todavia merecem destaque os trabalhos de Lara (1993, 2001, 2004a, 2004b, 2006, 2007, 2008, 2009, 2011); Tálamo e Lara (2009); Maimone e Tálamo (2011), que seguiram os fundamentos da França no que diz respeito à linguagem documentária, mas buscaram nos seguimentos espanhóis caminhos para tratar dos aspectos linguísticos. Lara (1993, 2001, 2004a, 2004b, 2006, 2007, 2008, 2009, 2011),

por exemplo, trabalha em busca de consolidar a linguística documental no Brasil seguindo os passos de García Gutiérrez (1984) e Izquierdo Arroyo (1990, 1993, 1995).

A terminologia, por sua vez, tão importante como fundamento e parâmetro para elaboração de tesouros e consolidação de uma linguagem especializada tanto para a representação do conhecimento como para a representação da informação, surge na década de 1930, oriunda do Círculo de Viena, logo fundamentada na lógica, apresentou seus segmentos, em destaque, nos trabalhos encabeçados pelo engenheiro austríaco E. Wüster, quem deu origem a teoria geral da terminologia (TGT) (Gomes, 2021; Almeida, 2003, 2006).

Embora estivesse diante de problemas semânticos, cognitivos e científicos, já que era preciso sistematizar um conjunto de termos que pudessem estabelecer a relação semântica, a conexão intelectual entre linguagem e realidade, além de funcionar como recurso para promover de modo eficiente a comunicação científica se manteve no plano da forma, em que o sentido é decorrente da construção social ou na relação sintática.

Por sua vez, o avanço nos estudos sobre a terminologia proposto por Cabré (1993, 2003) encontra uma versão voltada para aspectos sociais. Outrossim, Cabré (2003) defende que a terminologia é um conjunto de expressões aplicada para evidenciar o pensamento especializado e, desse modo, transferir, comunicar e organizar. Para Almeida (2003, 2006), o estudo da terminologia é da preocupação da linguística, por sua vez da lexicologia e da lexicografia.

Diante disso, embora Cabré (1993, 2003) apresente um modelo social, com preocupações cognitivas, ainda é possível reconhecer que sua concepção de linguagem se restringe ao estudo da forma, já que a especialista defende que a terminologia é constituída por três áreas do conhecimento: a linguística, que se responsabiliza pelo estudo da forma; a filosofia, que se preocupa com os aspectos do conceito; e a ciências de especialidade, que oferece o pensamento e organização para que a terminologia possa efetuar a composição do vocabulário.

Isso demonstra que essa teoria da terminologia não reconhece que haja fundamentos da metafísica e da ontologia, ou seja, princípios filosóficos no processo de representação e mediação da realidade intrínsecos aos estudos da linguística e da filosofia da linguagem. Importante salientar que os estudos sobre a linguagem já implementaram aspectos teóricos que atrelam a linguagem aos fundamentos filosóficos, ou seja, o estudo da

linguagem não se resume somente o processo de descrição da realidade e de comunicação.

Frente a isso, fica evidente que os estudos norteadores do tratamento da linguagem e do documento na ciência da informação ainda estão apegados a teorias antigas ou, ao menos, que também apresentam concepções antigas acerca dos estudos da linguagem, por isso a preocupação com a linguagem é secundária e restrita às discussões sobre a forma, assim não há estratégias para explorar a interpretação do conteúdo através da linguagem.

Seguindo o critério de organização aplicado a esta pesquisa, a segunda linha da linguística no cenário mundial é de natureza russa. Estudos sobre a leitura e a interpretação remontam as obras apresentadas por V. Propp, em que há uma relação com o folclore russo e técnicas de generalidade para observar padrões nas histórias e, assim, poder classificá-las. É bem verdade que essa linha de análise inspirou A. J. Greimas (1978) no desenvolvimento de seu percurso gerativo de sentido.

Outro teórico russo, com passagens pelos principais círculos de desenvolvimento da linguística, é R. Jakobson. Além de explorar aspectos da semiótica, como a ciência do signo, explorando tanto Saussure, como também apresentando interesse pela produção filosófica de C. S. Peirce. No entanto, o cerne do seu trabalho se direciona às preocupações com a comunicação, assim decorrente da teoria matemática da informação, apresenta sua teoria da comunicação, ampliando os atributos dos elementos da comunicação, enfatizando o código e o conteúdo da mensagem, além de destacar o papel importante do receptor na interação da comunicação. Nessa linha, também direcionou seus estudos à informação literária e se preocupou com os recursos da linguagem responsáveis pela evidência do conteúdo poético (Nöth, 1996; Fiorin, 2007; Mussalim & Bentez, 2011). Com isso, desdobrou os elementos da comunicação em funções da linguagem, em que se ocupou da complexidade da função poética quando comparada à função referencial (Nöth, 1996; Fiorin, 2007; Mussalim & Bentez, 2011).

Outrossim, o ponto mais relevante para o desenvolvimento da linguística mundial foram os desdobramentos do grupo denominado círculo de Bakhtin (Ponzio, 2011). M. Bakhtin e seus discípulos inicialmente se dedicaram ao estudo do signo, um viés do signo linguístico de Saussure, porém com acréscimos marxistas, assim apresentando uma versão de signo ideológico, em que se evidencia o significado e a significação

gerados em uma arena discursiva (Bakhtin, 2006, 2011; Ponzio, 2011).

Nessa linha, Bakhtin e seus discípulos debatem o papel do sujeito discursivo e a importância da alteridade na relação entre sujeitos inseridos no mesmo cenário discursivo e, muitas vezes, com interesses distintos acerca do objeto de discussão. Desse modo, não é objeto que motiva a tomada de decisões, mas sim a ideologia que se manifesta no objeto através de um signo (Bakhtin, 2006, 2011; Ponzio, 2011).

Além disso, Bakhtin apresenta a teoria do gênero textual, revolucionária em muitos aspectos, principalmente no que tange à concepção de texto, sua formatação estrutural, sua organização física e formal, bem como sua função social na transmissão do conteúdo, ou seja, os gêneros textuais fomentam pressupostos na comunicação, já que estabelecem regras de formatação do texto a fim de evidenciar as intenções do texto e modalizações de sua manifestação nos meios sociais adequados e ajustados aos diversos gêneros textuais (Bakhtin, 2006, 2011; Ponzio, 2011).

Os estudos russos acerca da linguagem sempre estão preocupados com a cultura e com a presença do outro na interação comunicacional. Derivado de uma herança marxista, esses estudos não se apegam apenas à forma, ou se detém aos aspectos do conteúdo como se fossem particularidades que se unem para formar uma peça maior. Ao contrário, os russos comprovam que os aspectos linguísticos são complexos e seus elementos não são explorados isoladamente, mas sim apenas funcionam em conjunto, portanto a linguagem é uma somatória de fenômenos que visam ao discurso, à comunicação e, principalmente, às pessoas (Ponzio, 2011).

Ainda nessa linha cultural, destaca-se o trabalho de Lotman (1979) à frente da chamada semiótica da cultura. Apresentando uma base fundamentada na semiologia de Saussure, sua aplicabilidade apresenta avanços inerentes aos fatores ideológicos e culturais, sendo estes determinantes para a fomentação do significado e norteadores da interpretação acerca da significação. Nessa perspectiva, Lotman (1979) apresenta sua concepção de semiosfera, ou seja, uma camada social em que uma espécie de ecossistema à base do signo orientasse a conduta dos humanos, ou seja, o pensamento humano é consequência dos produtos semânticos e semióticos construídos pelas pessoas pertencentes a uma semiosfera determinada (Schnaiderman, 1979).

Outro estudo que merece atenção, embora não seja voltado especificadamente para a linguística, utiliza-se da linguagem, ou melhor, do signo para compreender o processo cognitivo. Trata-se

da teoria da linguagem e da mente, também conhecido como sociointeracionismo, ou construtivismo histórico, desenvolvida por L. S. Vygotsky. Uma teoria da psicologia, aplicada à aprendizagem e à cognição, em que prevê a importância da experiência, da relação social e histórica e da linguagem como condutora de significados. Acionada pela zona de desenvolvimento proximal, a pessoa em um estágio potencial de aprendizagem pode acessar novos conhecimentos frente à mediação de pessoas mais experientes, portanto a interação é fundamental no processo de significação (Vygotsky, 2007, 2008).

Frente a essa exposição teórica, observa-se que essa corrente linguística se direciona a estudos cuja finalidade é buscar respostas sobre a significação. Entretanto, a concepção do significado não é previamente estabelecida como herança social, ao contrário, o sentido é construído em parcimônia, já que a alteridade é ponto fundamental e, assim, a significação se constrói em parceria, analisando a compreensão do outro, possibilitando a participação não apenas daqueles que fazem parte de grupos de interesses comuns, mas sim garantindo o acesso de todos à informação. É fato que a defesa da tese russa foi mais fácil do que transformá-la em realidade, porém suas bases teóricas podem contribuir para o desenvolvimento de práticas, estratégias, métodos de análise e interpretação de texto.

Como já foi dito, estudiosos de Brasil e da Espanha já olharam para as teorias da linguística argumentativa e textual, todavia os estudos russos podem ampliar as discussões, visto a preocupação com os gêneros textuais e com os gêneros discursivos.

A indexação é conduzida por protocolos, normas e políticas que garantem a convergência da linguagem natural para a linguagem técnica, a terminologia, bem como o uso desta para a confecção de tesouros, os quais são inseridos nos sistemas informacionais para facilitar a interoperabilidade e navegação, contudo, antes da escolha dos conceitos que representam o documento, é preciso realizar uma leitura, a qual se estabelece pelo teor técnico-científico de cunho especializado, portanto inserido em uma semiosfera específica, em que é preciso questionar o teor ideológico do signo que se converte em terminologia a fim de se garantir que a prática de indexação não valorize apenas o discurso padrão. Desse modo, a prática de indexação pode questionar os valores opressores, recuperar conceitos inerentes aos grupos excluídos, principalmente quando se trata de indexação de documentos em ciências humanas.

Por fim, conforme a esquematização proposta, a terceira corrente da linguística em âmbito mundial, trata-se da corrente estadunidense. Ela surge com os estudos formalistas, em destaque o trabalho de Bloomfield (1973), que dialoga com fundamentos saussureanos. Porém, um grupo de filósofos da linguagem acabam dedicando seus estudos para compreender a relação língua, linguagem e mente, assim transitando de uma abordagem formalista e ultrapassada para uma abordagem cognitivista e moderna. Encabeçando esses estudos estão pesquisadores do MIT, destacando N. Chomsky.

Nessa perspectiva, os estudos da linguagem se unem pelo viés da linguística e da filosofia da linguagem, do mesmo modo que muitos dos resultados contribuem para o desenvolvimento das duas áreas. De modo geral, pode-se apontar os estudos da lógica na compreensão da sentença, ou seja, inicialmente os estudos de N. Chomsky exploraram a formatação sintática da língua a fim de compreender os fundamentos do sistema computacional de uma língua específica e como esta poderia interferir na formatação de uma mente computacional humana.

Por sua vez, Chomsky dá seguimento ao que ficou conhecida como gramática gerativa. Com avanços de seus estudos, Chomsky passa a defender uma composição biológica humana capaz de realizar a cognição, dessa maneira essa composição seria responsável por possibilitar uma gramática universal, da qual seria decorrente a capacidade humana de desenvolvimento dos aspectos linguístico, denominado de Língua-I a partir do *input* promovido pela interação com a língua social, denominada de Língua-E (Chomsky, 2009, 2014; Kenedy, 2016).

Com as preocupações voltadas para fatores inerentes à composição da mente, ou melhor, para a composição da cognição, os estudos de Chomsky (2009, 2014) se tornaram fundamentais para o desenvolvimento de outras linhas da linguística, como a psicolinguística e os estudos de aquisição da linguagem, além da neurolinguística, estudos que exploram os recursos da cognição e qual é o papel da linguagem na ativação desses recursos. Ademais, os estudos sobre a psicologia da linguagem foram fundamentais também para o desenvolvimento da semântica cognitiva e da linguística computacional.

Além de Chomsky, o MIT proporcionou outros estudos importantes sobre a mente computacional, a máquina semântica e os recursos de compreensão da realidade decorrentes de uma interação linguagem, cognição e referência. Nessa linha, destacam-se os trabalhos de Dowty (1979), Fodor (1984, 2014), Vendler (1984), Jackendoff

(1985, 1990), Pinker (2002) e Talmy (2017), entre outros, que debatem sobre a natureza da linguagem na mente humana, bem como os aspectos de significação cognitiva, ou seja, se a significação é decorrente do sentido psicológico, ou se a significação é motivada pelos fatores extralinguísticos inerentes à natureza do objeto, do fenômeno ou do evento.

De modo geral, poucos são os estudos dessa linha que ganharam fama e destaque entre a comunidade linguística no Brasil e na Espanha, em destaque essas discussões se concentram nos trabalhos de N. Chomsky, no entanto muitos especialistas no Brasil têm explorado suas fundamentações principalmente pela crescente discussão sobre neurociência e o impacto dos fatores mentais e psicológicos para a cognição humana. Desse modo, trabalhos que exploram a mente computacional e a linguística computacional ganham destaque no cenário brasileiro, em áreas que envolvem a computação, já sendo possível reconhecer algumas abordagens no universo da ciência da informação, através do processamento em linguagem natural.

Embora não sejam teorias recentes, os estudos sobre organização e representação do conhecimento e da informação não estão explorando essa linha da linguística com profundidade. Na Espanha, trabalhos de Gil-Leiva apresentam aproximação com fatores cognitivos, baseando-se nos estudos de Van Dijk sobre a linguística discursiva e textual. Porém, o trabalho de Van Dijk ou até mesmo de Ducrot não aprofundam a análise nas conjunturas da neurociência.

Os trabalhos dessa terceira linha da linguística poderiam contribuir para analisar os fatores psicológicos que envolvem o trabalho de leitura, tratamento e análise do documento, compreendendo os caminhos mentais para a compreensão dos temas, para a identificação dos conceitos e para a seleção dos termos aplicados na indexação.

Além disso, abre-se a hipótese de discutir sobre uma máquina semântica para a indexação fundamentada no *input* de tesouros que poderiam filtrar a linguagem natural durante a busca e a recuperação do documento. Outrossim, é possível haver a contribuição para a indexação automática, frente à compreensão da mente computacional e avançar no entendimento do sistema informacional das línguas, além de direcionar mecanismos para estabilizar a tradução envolvendo tesouros em diferentes línguas e ainda sistematizar os avanços da psicolinguística e da linguística computacional.

De modo geral, tanto as teorias da linguística quanto as teorias da filosofia da linguagem apresentam um cenário extremamente amplo e que vem abordando vários fundamentos que envolvem a linguagem em direção ao sentido, ao significado e à significação, assim relacionando seus atributos à forma, aos recursos para que a forma gere significação e acesso ao conteúdo, seja pelo enunciado, pela enunciação, pela semiótica, pelo discurso, utilizando-se dos elementos da linguagem, como também avançando nos quesitos que envolvem a cognição humana.

4. Linguagem e indexação

A indexação é um processo fundamentado em sistematizações que visam a uma espécie de padrão para que os mecanismos de busca sejam eficientes na recuperação da informação. Consequentemente, o sistema em que as informações são inseridas são organizados através da lógica. Somado a isso, a prática do profissional da indexação é orientada por políticas e normas, bem como a linguagem adequada para a sua realização, ou seja, uma linguagem documental elaborada a partir de terminologia especializada no conhecimento científico ou cultural que será conteúdo da indexação.

Assim, é válido discutir as marcações linguísticas na organização do enunciado de um texto e reconhecê-las para direcionar as estratégias de leitura e interpretação. Com isso, propõe-se analisar a presença de modais presentes nas estratégias de indexação.

Nessa linha, a indexação se consolida através de uma leitura especializada, denominada de leitura documental, em que ocorre o tratamento da informação presente no documento analisado a fim de que os conceitos sejam evidenciados e, consequentemente, possam ser traduzidos em uma linguagem de indexação, ou seja, esses conceitos são nomeados por termos que perfazem o banco de dados do sistema informacional. Diante disso, a indexação visa a um procedimento realizado com exatidão, ou, ao menos, sob perspectivas que ofereçam ao sistema informacional mecanismos precisos para que a indexação execute sua função com eficiência.

Sob essas determinações e toda a tradição pressuposta no processo de indexação, não haveria mais problemas a serem questionados e refletidos, já que, como indicado, a indexação está fundamentada em procedimentos lógicos e protocolares, assim bastaria que os protocolos e a política de indexação conduzissem o trabalho do indexador e a formulação do sistema informacional. Nessa linha, o usuário padrão, conhecedor

dos pressupostos protocolares, da política aplicada àquele sistema de informação e conhecedor da linguagem de indexação realizaria o percurso de acesso com tranquilidade.

Porém, o que se evidencia é que o processo de indexação apresenta limites, ou melhor, se por um lado há requisitos para funcionar com eficiência em sistemas informacionais fechados; por outro, diante de sistemas informacionais mais complexos e estes submetidos a usuários de diversos perfis, as regras de indexação se deparam com muitas variáveis, embora haja controle de exaustividade, especificidade, revocação e de precisão.

Como já detectado, o ponto frágil desse processo é a linguagem, pois com o uso e aplicação da linguagem também se estruturam o contexto, a intencionalidade, a aspectualidade de figuras e temas, além de fatores cognitivos inerentes à experiência, à percepção dos fatores que se apresentam diante do usuário no momento de efetuar a busca e ao conhecimento da especificidade do assunto e dos temas abordados, assim como conhecimento linguístico para nomear seus objetos, pensamentos e resultados.

A indexação é parte importante dos serviços de organização e representação da informação, por sua vez, para alcançar seu resultado final, a informação é tratada, analisada e significada em várias camadas e níveis, por isso os estudos que correlacionam a indexação e as teorias da linguagem, sejam pela linguística, seja pela filosofia da linguagem devem atuar com cuidado e atenção às propostas das teorias, com o fito de que se concentre em uma discussão filosófica que envolva a significação, o enunciado e a enunciação, pois a indexação se posiciona ancorada ao enunciado e dele extrai uma parte mínima de significação, aquilo que se julga suficiente para representar toda a enunciação e todo os fatores interdiscursivos presentes no documento.

Evidentemente que a indexação não apresenta como preocupação central fatores relacionados à significação apresentada no texto, pois seu objetivo se restringe a representá-lo em um sistema informacional e que seus mecanismos sejam funcionais para a recuperação desse documento, o qual será ressignificado posteriormente pelo seu leitor/usuário.

No entanto, a questão que fica diante desses fatores é sobre o roteiro de indexação, ou seja, quais questionamentos o indexador deve fazer ao texto para extrair a informação? O modelo midiático fundamentado nos pronomes modais de interrogação (Quem? O quê? Onde? Quando? Como? Por quê?), ou ainda as estratégias à base

dos nocionais adverbiais se mostram eficientes diante da nova demanda?

Os sistemas de informação com acesso aberto à pluralidade de usuário têm condicionado os mecanismos de busca, os quais recuperam músicas apenas com o cantarolar da melodia, ou trechos dos textos. Assim, reconhece-se que as buscas se perfazem por novos recursos modais que se misturam aos tradicionais. Com isso, é recorrente a busca bibliográfica, a qual é conduzida pelo nome do autor ou da obra. Desse modo, os modais tradicionais podem trazer as respostas com tranquilidade.

Além disso, as pesquisas podem ser realizadas por assunto, data da publicação, tema ou recorte temático. Diante disso, essas buscas recorrem a aspectos semânticos, por isso indexação à base de tesouro proporciona resultados satisfatórios, já que há uma correlação hierárquica, associativa e partitiva, ou seja, faz-se uso da semântica lexical explorando a lexicologia e seus correspondentes em hiperônimo x hipônimo; meronímia e sinonímia. Ademais, muitos sistemas de informação fundamentados na lógica realizam a varredura e a recuperação através de mecanismos booleanos.

Por seu turno, o tipo de pesquisa que promove dificuldades de recuperação é a busca fundamentada no arcabouço temático, meio de investigação que o usuário pode realizar uma busca constituindo um hipertexto, ao ainda uma busca que se realiza através de aspectos transversais. Esse modelo de pesquisa é fundamentado em um determinado assunto ou recorte temático, sob o qual se estrutura a construção discursiva sobre o item em diversos campos científicos e culturais, ou seja, a pesquisa não se realiza sob o aprofundamento do assunto ou tema em um determinado referente a um determinado autor, um determinado segmento ou um determinado campo científico, mas sim fundamentado em recuperar o que for possível de produção sobre o objeto investigado.

Diversas produções sociais e culturais proporcionam a intertextualidade e a interdiscursividade. Com isso, o trabalho do profissional mediador da informação fica mais complexo, pois recuperar a informação tem sido um trabalho mais amplo e de diálogos mais profundos. Sendo assim, reconhece-se que além da divergência referente à aplicação do estrato linguístico para efetuar a pesquisa, há também fatores que impactam na abrangência da pesquisa, pois envolve significação, enunciação, intencionalidade e discurso.

Sob esses aspectos atrelados ao processo informacional e à qualidade da informação recuperada, seja sobre a profundidade do seu conteúdo,

seja sobre a precisão referente ao que se busca, fica evidente que novas demandas estão condicionando os mecanismos de tratamento informacional, tanto no que tange à representação, quanto ao que se refere à recuperação, à mediação, à compreensão, à aplicação e ao uso.

Sendo assim, é factual que o impacto da linguagem nesse recorte simbólico da interação informacional, por isso a necessidade de ampliação de estudos que explorem as diversas teorias da linguagem, da linguística e da filosofia da linguagem para estabelecer critérios de leitura e interpretação no tratamento da informação no reconhecimento dos conceitos. Esses conceitos são estruturados sob qual estratégia e para ser recuperado por quais grupos? Pois fica evidente que a nova demanda informacional exige novos modelos modais. Modelos estes que não se limitem ao enunciado tampouco ao conjunto de conceitos que circundam a superfície do texto. Esses modais são reflexo das formas de vida que orbitam o documento.

5. Considerações finais

A informação é um dos mais importantes recursos humanos para compreender a si e à realidade, assim exerce um papel extremamente importante na construção do conhecimento. Todo o percurso que envolve o tratamento da informação, a sua representação e organização em sistemas informacionais é conduzido pela linguagem. Essa linguagem se evidencia por ser uma linguagem especializada, elabora sob um estrato linguístico específico que é a terminologia, o qual é responsável por condensar pensamentos, conhecimentos, comunicações e ativar o processo cognitivo em seus usuários a fim de que estes compreendam sua atuação em determinados contextos e funções.

Assim sendo, esta pesquisa evidenciou que a terminologia é o principal recurso para a elaboração de linguagem documental, por conseguinte de tesouros, porém o que se apurou é que a maior parte dos fundamentos linguísticos que assessoram a indexação estão ancorados a uma teoria estruturalista, embora tantas outras teorias da linguagem, da linguística e da filosofia da linguagem já apresentaram novas abordagens e evoluções importantes para o tratamento da significação.

Outro ponto importante debatido é o fato de que a indexação é um processo que se realiza sob concepções protocolares a fim de que a prática da indexação seja amparada por uma política de indexação, por um sistema informacional à base da lógica e por uma linguagem de indexação que garanta precisão na recuperação da informação,

por isso a indexação se configura como uma etapa bastante exata no cenário da organização da informação.

Ademais, constatou-se que sendo a linguagem principal insumo para o processo de indexação, embora as normas e os protocolos fundamentem um conjunto de linguagem especializada para a representação do documento, a terminologia é um estrato linguístico e, por isso, sobre impactos de variações semânticas quando inserida no discurso que se expressam na composição e comunicação do documento. Com isso, não se pode ignorar o impacto da linguagem nesse processo. Outrossim, sendo evidente que o processo de organização e representação não se finaliza, quando o documento é cadastrado no banco de dados do sistema informacional, isto é, a linguagem é gestora das estratégias modais e cognitivas da recuperação da informação.

No entanto, também se evidenciou que com o advento da tecnologia e a ampliação do acesso à informação através de sistemas informacionais amplos, abertos e gerais e, além disso, acessado por usuários de perfis diversificados, manter a precisão e a eficácia na recuperação da informação tem se tornado um grande desafio para os sistemas informacionais e para os profissionais responsáveis pela elaboração dos objetos informacionais que conduziram o acesso à informação.

Ademais, esta pesquisa também evidenciou o impacto dos fatores cognitivos na leitura, interpretação e busca pela informação e a interação com o documento, seja pelo indexador, seja pelo leitor e apontou a necessidade de se refletir sobre as estratégias de indexação referentes às estratégias modais. A tradição inerente à indexação viabilizou entre tantos modelos os pronomes interrogativos, os aspectos nocionais, facetas, todavia discutiu-se o fato de que os aspectos modais conduzem o formato da indexação, e a estratégia por trás da formatação dos modais no tratamento da informação é o modelo que impacta na conduta do usuário ao realizar suas buscas.

Agradecimentos

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

Referências

- Almeida, Gladis. M. B. (2003). O percurso da terminologia: de atividade prática à consolidação de uma disciplina autônoma. // TradTerm. 9:1, 211-222.
- Almeida, Gladis. M. B. (2006). A teoria comunicativa da terminologia e a sua prática. // Alfa. 50:2, 85-101. São Paulo.

- Bakhtin, Mikhail. M. (2006). Marxismo e filosofia da linguagem: problemas fundamentais do método sociológico da linguagem. Trad. Michel Lahud, Yara Frateschi Vieira. 12. São Paulo: Hucitec.
- Bakhtin, Mikhail. M. (2011). Estética da criação verbal. Trad. Paulo Bezerra. 6. São Paulo: Editora WMF Martins Fontes.
- Bloomfield, Leonard. (1973). Language. London: George Allen & Unwin LTD.
- Cabré, Maria. T. (1993). La terminología: teoría, metodología, aplicaciones. Trad. Carles Tebé. Barcelona: Editoria Antàrtaida/Empúres.
- Cabré, Maria. T. (2003). Theories of terminology. // Terminology. 9:2. 163-199.
- Chomsky, Noam (2009). Linguagem e mente. Trad. Roberto Leal Ferreira. São Paulo: Editora UNESP.
- Chomsky, Noam (2014). A ciência da linguagem: conversas com James McGilvray. Trad. Gabriel Ávila Othero, Luisandro Mendes Souza, Sérgio de Moura Menuzzi. São Paulo: Editora Unesp.
- Corrêa, Renato. F. e Lapa, Remi C. (2015). Panorama de estudos sobre Indexação Automática no âmbito da Ciência da Informação no Brasil (1973-2012). // Ciência Da Informação. 42:2. <https://doi.org/10.18225/ci.inf.v42i2.1385>
- Dowty, David. R. (1979). Word meaning and Montague grammar: The Semantics of verbs and times in Gerative Semantics and in Montague's PTQ. Dordrecht: Kluwer Academic Publishers.
- Fiorin, José. L. (2007). Introdução à lingüística. 5. São Paulo: Contexto.
- Floch, Jean. M. (1985) Petites mythologies de l'oeil et de l'esprit: pour une sémiotique plastique. Amsterdam: Hadès-Benjamins.
- Fontanille, Jacques (2012). Semiótica y literatura: ensayos de método. Trad. Desidero Blanco. Lima: Universidad de Lima, Fondo Editorial.
- Fodor, Jerry. A. (1984) El lenguaje del pensamiento. Trad. Jesús Fernández Zulaica. Madrid: Alianza Editorial, SA.
- Fodor, Jerry. A. (2014). Mente e linguaggio. Trad. Francesco Ferretti. Bari: Editori Laterza.
- Fujita, Mariangela. S. L. (2016). A linguagem documentária na negociação de uma política de indexação para bibliotecas universitárias: procedimentos e estratégias da pesquisa-ação integral. // Revista Conhecimento em Ação. 1:1, 2-17. doi:<https://doi.org/10.47681/rca.v1i1.3555>. Rio de Janeiro.
- Fujita, Mariangela. S. L. (2021). A política de indexação para representação e recuperação da informação. // Gil Leiva, Isidoro, Fujita, Mariangela. S. L. (eds). // Política de indexação [online]. São Paulo: Cultura Acadêmica; Marília: Oficina Universitária, 2021. 257 p. ISBN: 978-65-5954-130-0. <https://doi.org/10.7476/9786559541300>.
- Fujita, Mariangela. S. L.; Isidoro. Gil-Leiva.(2009). Políticas de indexação na américa latina. // Ibersid: revista de sistemas de información y documentación. 3:1, 155-162. ISSN 2174-081x; issn 1888-0967. <http://hdl.handle.net/20.500.11959/brapci/167288>.
- Fujita, Mariangela. S. L.; Gil-Leiva, Isidoro. (2014). Avaliação da indexação por meio da recuperação da informação. // Ci. Inf. 41:1, 50-66. Brasília.
- García Gutiérrez, Antonio (1984). Lingüística documental: aplicación a la documentación de la comunicación social. Barcelona: Mitre.
- Gil-Leiva, Isidoro (2021). Aspectos conceituais da Indexação. Gil Leiva, Isidoro, Fujita, Mariangela. S. L. (eds). // Política de indexação [online]. São Paulo: Cultura Acadêmica; Marília: Oficina Universitária, 2021, 257 p. ISBN: 978-65-5954-130-0. <https://doi.org/10.7476/9786559541300>

- Gil-Leiva, Isidoro; Rodríguez Muñoz, José Vicente; (1996). El procesamiento del Lenguaje natural aplicado al análisis del contenido de los documentos. // Revista General de Información y Documentación, 6.2. Madrid: Servicio de Publicaciones Universidad Complutense.
- Gomes, Hagar. E. (2021). Terminologia e estrutura conceitual. // PontodeAcesso, 15:3. <https://doi.org/10.9771/rpa.v15i3.47464>
- Izquierdo Alonso, Monica; Moreno Fernández, Luis Miguel. (2010). Perspectives of the studies on document abstracting: towards an integrated view of models and theoretical approaches. // Journal of Documentation. 66:4, 563-584.
- Izquierdo Alonso, Monica; Moreno Fernández, Luis Miguel; Sánchez Domínguez, M del Carmen (2012). Reflexiones sobre el resumen documental y la práctica resumidora: revisión y sistematización de enfoques teórico-prácticos. // Actas del X Congreso ISKO Capítulo Español (Ferrol, 2011), Universidade da Coruña (España).. 353-369.
- Izquierdo Arroyo, José María (1990). Esquemas de Lingüística Documental. Barcelona: PPU. 3 vols.
- Izquierdo Arroyo, José María (1995). La organización documental del conocimiento. Madrid:Tecnidoc.
- Izquierdo Arroyo, José María (1993). De la Semiótica del Discurso a la Semiótica Documental. Epílogo. // Moreiro González, J. A. Aplicación de las Ciencias del Texto al resumen documental. Madrid: Univ. Carlos III; // Boletín Oficial del Estado. 199-216.
- Jackendoff, Ray S. (1985). Semantics and Cognition. Massachusetts: The MIT Press.
- Jackendoff, Ray S. (1990) Semantics Structures. London: The MIT Press.
- Kenedy, Eduardo. (2016). Curso básico de linguística gerativa. São Paulo: Contexto.
- Lara, Marilda L. L. G. (1993). Algumas contribuições da semiologia e da semiótica para a análise das linguagens documentárias. // Ciência da Informação. 22:3. <https://doi.org/10.18225/ci.inf.v22i3.480>
- Lara, Marilda L. L. G. (2001). O unicórnio (o rinoceronte, o ornitorrinco...), a análise documentária e a linguagem documentária. // DataGramZero. 2:6. <http://hdl.handle.net/20.500.11959/brapci/5294>.
- Lara, Marilda. L. L. G. (2004a) Linguagem documentária e terminologia. // Transinformação, 16:3, 231-240. <https://doi.org/10.1590/S0103-37862004000300003>.
- Lara, Marilda. L. L. G. (2004b) Diferenças conceituais sobre termos e definições e implicações na organização da linguagem documentária. // Ciência da Informação. 33:2. <https://doi.org/10.18225/ci.inf.v33i2.1050>.
- Lara, Marilda L. L. G. (2006). É possível falar em signo e semiose documentária?. // Encontros Bibli: Revista Eletrônica de Biblioteconomia e Ciência da Informação. 18-29. <https://doi.org/10.5007/1518-2924.2006v11nesp3p18>
- Lara, Marilda L. L. G. (2008). Informação, informatividade e linguística documentária: alguns paralelos com as reflexões de Hjørland e Capurro. // DataGramZero. 9:6. <http://hdl.handle.net/20.500.11959/brapci/6550>.
- Lara, Marilda L. L. G. (2009). Linguística Documentária: seleção de conceitos. 2009. Tese (Livre Docência em Análise Documentária) - Escola de Comunicações e Artes, Universidade de São Paulo, São Paulo. <https://doi.org/10.11606/T.27.2019.tde-21112019-191517>
- Lara, Marilda L. L. G. (2011). Conceitos de organização e representação do conhecimento na ótica das reflexões do grupo temma. // Informação & Informação. 16:2, 92-121. <https://doi.org/10.5433/1981-8920.2011v16n2p92>
- Lopes, Ivan. C.; Hernandes, N. (2021). Semiótica: objetos e práticas. 2. São Paulo: Contexto.
- Lotman, Iurii. M. (1979). Sobre o problema da tipologia da cultura. Semiótica Russa. Schnaiderman, Boris (org.). São Paulo: Perspectiva. 31-42.
- Maimone, Giovana D.; Tálamo, Maria de Fátima G. M. (2011). Linguística e terminologia: contribuições para a elaboração de tesouros em ciência da informação. // DataGramZero. 12:2. <http://hdl.handle.net/20.500.11959/brapci/7356>
- Milani, Sebastião. E. (2016). Relato da obra de Ferdinand de Saussure. Rio de Janeiro: Barra Livros.
- Mussalim, Fernanda.; Bentes, Anna Chistina. (2011). Introdução à linguística: fundamentos epistemológicos, 3:5. São Paulo: Cortez.
- Mussalim, Fernanda.; Bentes, Anna Chistina. (2012). Introdução à linguística: domínios e fronteiras, 1:9. São Paulo: Cortez.
- Nöth, Winfried. (1996). A semiótica no século XX. São Paulo: ANNABLUME.
- Nöth, Winfried.; Santaella, Lúcia. (2017). Introdução à semiótica: passo a passo para compreender os signos e a significação. São Paulo: Paulus.
- Pinker, Steven. (2002). O instinto da linguagem: como a mente cria a linguagem. Trad. Cláudia Berliner. São Paulo: Martins Fontes.
- Ponzio, Augusto (2011). A revolução bakhtiniana: o pensamento de Bahktin e a ideologia contemporânea. Trad. Valdemir Miotello. São Paulo: Contexto.
- Schnaiderman, Boris. (1979). Semiótica Russa. Trad. Aurora Fomoni Bernardini, Boris Schnaiderman, Luci Seki. São Paulo: editora Perspectiva.
- Saussure, Ferdinand. (2006). Curso de linguística geral. Trad. Antônio Chelini, José Paulo Paes, Isidoro Blikstein. 27. São Paulo: Cultrix.
- Tálamo, Maria de Fátima. G. M.; Lara, Marilda. L. L. G. (2009). Interface entre linguística, terminologia e documentação. // Brazilian Journal of Information Science. 3:2. <https://doi.org/10.5016/brajis.v3i2.361>
- Talmy, Leonard (2017). The targeting system of language. Massachusetts: The MIT Press.
- Unesco. (1981). Indexing principles. Trad. Maria Cristina Mello Ferreira Pinto. Belo Horizonte: R. Esc. Bibliotecon.
- Vender, Zeno (1984) The matter of minds. New York: Oxford Clarendon Press.
- Vigotsky, Lev. S. (2007). A formação social da mente: o desenvolvimento dos processos psicológicos superiores. Trad. José Cipolla Neto, Luís Silveira Menna Barreto, Solange Castro Afeche. 7. São Paulo: Martins Fontes.
- Vigotsky, Lev. S. (2008). Pensamento e linguagem. Trad. Jefferson Luiz Camargo. 4. São Paulo: Martins Fontes.
- Zilberberg, Claude. (2006). Razão e poética do sentido. Trad. Ivã Carlos Lopes, Luiz Tatit, Waldir Bevidas. São Paulo: Editora da Universidade de São Paulo.
- Zilberberg, Claude. (2015). La estructura tensiva. Trad. Desidero Blanco. Universidad de Lima: Fondo Editorial.

Enviado: 2023-04-01. Segunda versión: 2023-09-13.
Aceptado: 2023-11-08.

Implementación de los repositorios de datos de investigación en las universidades públicas españolas: estado de la cuestión

Implementation of research data repositories in Spanish public universities: state of the art

**Francisco Javier MARTÍNEZ MÉNDEZ (1), Ana Alice BAPTISTA (2),
Rosana LÓPEZ CARREÑO (3), Ángel María DELGADO VÁZQUEZ (4)**

(1) (3) Facultad de Comunicación y Documentación de la Universidad de Murcia, javima@um.es, rosanalc@um.es

(2) Escola de Engenharia da Universidade do Minho, analice@dsi.uminho.pt

(4) Área de Biblioteconomía y Documentación, Universidad Pablo de Olavide, amdelvaz@gmail.com

Resumen

La Ciencia Abierta se abre paso poco a poco en el seno de las instituciones de educación superior. Aspira a conseguir que los resultados de las investigaciones científicas sean accesibles, reutilizables y transparentes. Para ello, es especialmente importante el acceso abierto a los datos de investigación. En este trabajo se analiza cómo se está llevando a cabo el depósito de los conjuntos de datos en las universidades públicas españolas, su nivel de implementación, qué plataformas se están implementando y conocer si se está orientando y apoyando a los investigadores en esta tarea. Los resultados obtenidos muestran una realidad diversa e incipiente. Por un lado, están dos consorcios regionales de universidades (Cataluña y Madrid) que han apostado claramente por la gestión de los conjuntos de datos de investigación. En el otro, aparece un modelo estándar de depósito de esta información como si de colecciones de un repositorio institucional se tratara. Las universidades con más conjuntos de datos publicados pertenecen a uno de los consorcios, implementan un software más avanzado y son un modelo a considerar por las otras instituciones, especialmente aquellas que apenas han publicado unos pocos conjuntos de datos. El apoyo de las administraciones regionales es también un factor a considerar y una oportunidad para las comunidades con varias universidades.

Palabras clave: Ciencia abierta. Conjuntos de datos de investigación. Repositorios de datos de investigación. Universidades públicas. España.

1. Introducción

En el transcurso de muchas investigaciones se generan datos que sirven para la obtención de resultados y la extracción de conclusiones. Dentro del campo de la Ciencia Abierta (UNESCO, 2021) viene cobrando impulso la idea de publicar, junto al artículo o informe derivado de la investigación el conjunto de datos empleados en la misma con vistas a favorecer la reutilización de los mismos por parte de otros investigadores, fomentando así a transparencia de la labor investigadora y garantizando la integridad de todo el

Abstract

Open Science is gradually making its way into the heart of higher education institutions. This paradigm shift aims to make the results of scientific research accessible and reusable by other researchers. In this regard, open access to research data is particularly important, in order to facilitate their reuse and increase their level of transparency. Our paper aims to analyse how the deposit of these datasets is being carried out in Spanish public universities, their level of implementation, which platforms are being implemented and whether researchers are being guided and supported in this task. The results obtained show a diverse and incipient reality. On the one hand, there are two regional consortia of universities (Catalonia and Madrid) that are clearly committed to the management of research datasets. On the other hand, there is a standard model for depositing this information as if they were collections in an institutional repository. Universities with more published datasets belong to one of the consortia, implement more advanced software and are a model to be considered by the others, especially those that have only published a few datasets. Support from regional administrations is also a factor to consider and an opportunity for communities with several universities.

Keywords: Open science. Research datasets. Research data repositories. Public universities. Spain.

proceso (Borghetti & Van Gulick, 2019). La Ciencia Abierta representa un nuevo paradigma que incorpora una visión holística del proceso de generación de conocimiento, a partir del diseño inicial de un proyecto y su desarrollo hasta la comunicación, difusión y preservación de sus resultados (Abadal et al, 2023). Dentro de este entorno cobra especial importancia la gestión de los datos de investigación: “un proceso diseñado para gestionar y difundir conjuntos de datos de alta calidad, que cumplan con los requisitos académicos, legales y éticos establecidos” (Alonso-Arévalo,

2019): Esta gestión se encuentra entre las principales líneas de trabajo, no ya futuras sino actuales, de las bibliotecas académicas (Federer & Qin, 2019), si bien no está implantada de manera estándar (Ayris & Ignat, 2018; Ashiq, 2022). Asimismo, esta gestión constituye uno de los ejes estratégicos de sobre los que se estructura la Estrategia Nacional de Ciencia Abierta de España (2023), recientemente aprobada.

El acceso a los datos de investigación es imprescindible para la reproducibilidad de los resultados científicos, facilita la cooperación interdisciplinar, estimula el crecimiento económico a través de mejores oportunidades para la innovación, permite la reutilización de datos, aumenta la eficiencia de los recursos, mejora la transparencia, la rendición de cuentas y la confianza en los resultados de la investigación científica (OCDE 2021). Esta gestión se lleva a cabo sobre los conjuntos de datos de investigación ('datasets') que recogen información estructurada y organizada recopilada o generada en el desarrollo de una investigación. Los conjuntos contienen información relevante y detallada que se utiliza para respaldar los objetivos, análisis y resultados del estudio científico. Los datos pueden ser primarios o secundarios, cuantitativos o cualitativos, y pueden estar disponibles en diferentes formatos y tamaños. Es totalmente recomendable que los datos cumplan los principios FAIR ('findability, accessibility, interoperability, reuse') que establecen pautas para la correcta gestión de datos (Wilkinson et al., 2016).

Con la idea de cumplir estos principios, se han desarrollado diversas recomendaciones de buenas prácticas para llevar a cabo su gestión, tales como las recogidas en el documento *FAIR Data Maturity Model. Specification and Guidelines* (2020) que ofrece un conjunto de directrices y una lista de comprobación relacionada con el cumplimiento de estos principios, o las que recogen Rocca-Serra et al. en su proyecto *FAIR Cookbook* (2023). En España, la red de bibliotecas universitarias elaboró su propio documento de recomendaciones para gestionar los datos de investigación (REBIUN, 2017). Los requerimientos técnicos, tecnológicos, así como la implicación bibliotecaria y de los responsables de la publicación en los repositorios, son una pieza fundamental para lograr una óptima implantación y, por tanto, para la visibilidad y reutilización de dichos datos en futuras investigaciones.

La implementación de los repositorios de datos de investigación en universidades y centros de investigación es un signo más de esta tendencia conducente a disponer en acceso abierto fuentes de información específicas y bien definidas en torno a los datos generados durante las investigaciones. Como se ha indicado anteriormente, esta

publicación debe seguir los principios FAIR en los propios datos, los mismos deben estar depositados en repositorios fiables y han de adoptar licencias tipo CC-BY o CC-0 (1) para compartirlos.

Debemos recordar que esto se debe llevar a cabo no solo para cumplir la solicitud de financiación pública de la investigación, algo ya habitual tanto en EE.UU. como en la Unión Europea (Redkina, 2019), sino para contribuir a una ciencia cada vez más abierta. Si estos requisitos no se cumplen del todo, disponiéndose estos conjuntos de datos de cualquier forma y formato (como ocurre con otros tipos documentales habituales de los repositorios institucionales), no se terminará de alcanzar uno de sus principales propósitos: la reutilización.

La responsabilidad de la gestión de datos recae, inicialmente, en los investigadores, si bien parece claro que van a necesitar apoyo para adaptarse a los nuevos requisitos específicos de esta gestión (Marín-Arraiza et al., 2019). En este punto en particular, las bibliotecas académicas tienen una nueva oportunidad (y un reto) de mejorar sus actuales servicios de apoyo a la investigación y aumentar su presencia y relevancia en el seno de las instituciones universitarias y de investigación (Angelozzi, S. M., 2020; Sheikh et al., 2023). En algunos casos se han constituido consorcios regionales, como en Cataluña y Madrid, para "acometer de forma colectiva los retos derivados de la Ciencia Abierta y procurar que su adopción se realice con el menor esfuerzo por parte de las universidades" (Alcalá y Anglada, 2019). Por ello, las políticas de acceso abierto de I+D+i que se están implantando han de indicar de manera explícita los procedimientos, estándares, formatos, licencias y lenguajes a adoptar para una normalización e identificación de los datos de investigación (EC, 2023). También es importante la revisión de los propios contenedores de los conjuntos de datos con el objeto de que estén en condiciones de ofrecer una respuesta útil en su identificación, descripción, catalogación clasificación y métricas de uso.

La integración de este tipo de contenidos no termina de llevarse a cabo en la gestión editorial de las revistas científicas donde se publican los artículos, donde apenas tienen cabida los conjuntos de datos. Esto ocurre a pesar de las iniciativas de grandes editoriales que implementan buscadores especializados en repositorios de datos, como Data Citation Index de Web of Science Group, Mendeley Data de Elsevier y Google Data Search. Este interés evidencia la dimensión del volumen de los conjuntos de datos de investigación generados (preceptiva o voluntariamente) y muestra claramente la necesidad de su localización y curación, como es el caso del proyecto

Data Curation Network (Johnston et al., 2017; Johnston et al., 2018).

Es innegable que los conjuntos de datos son una fuente de información en continuo crecimiento. Alcanzar su efectivo desarrollo es uno de los retos a los que se enfrenta el movimiento hacia la Ciencia Abierta (UNESCO, 2021; Bethencourt-Aguilar, 2022). El momento actual parece el más adecuado para reconsiderar el proceso de creación y mantenimiento, ayudando a establecer líneas y estrategias de gestión y depósito institucional que formen parte del signo distintivo de calidad científica propio de las instituciones de investigación. En España, la mayor parte de la investigación se lleva a cabo en universidades y centros públicos de i+d+i (Fundación CYD, 2023). Son estas instituciones las que deben procurar implementar estos repositorios de conjuntos de datos siguiendo las mejores prácticas posibles. Otra tarea importante es hacer partícipes a sus comunidades investigadoras de los beneficios de disponer en abierto los datos de sus investigaciones, especialmente por la transparencia de los trabajos de investigación (De Giusti, 2020). Este punto en particular ha cobrado una especial relevancia en los últimos meses por la proliferación de malas prácticas entre la comunidad científica, especialmente aquellas relacionadas con una hiperproducción científica incompatible con la calidad y la excelencia que se le presupone a la investigación.

Una vez expuesto el objeto de estudio, procede establecer el objetivo: analizar los repositorios de datos científicos de las universidades públicas españolas y establecer un diagnóstico de su nivel de implantación.

2. Metodología

Directamente relacionadas con la gestión de repositorios y datos de investigación están las organizaciones REBIUN y FECYT. En cuanto al total de universidades, en España, según los datos de Registro de Universidades, Centros y Títulos hay 86.

2.1. Universidades

De este total de universidades, 50 son públicas (47 presenciales, 1 no presencial y 2 especiales) y 36 privadas (31 presenciales y 5 no presenciales). La UNED es la universidad pública a distancia de gran implantación en todo el país y la Oberta de Catalunya, una universidad, pública, también de enseñanza a distancia, perteneciente al sistema universitario catalán y de gestión privada (si bien la consideramos como pública en el estudio). Las universidades “especiales” son la Menéndez Pelayo y la Internacional de Andalucía, especializadas en la organización de cursos de

verano y conferencias por lo que no desarrollan investigación y no van a ser objeto de estudio.

El total de universidades a estudiar lo van a formar las 47 universidades presenciales más las dos universidades a distancia citadas, en total serán 49 las instituciones a analizar.

Las universidades de la comunidad de Madrid (menos la Complutense a nivel de conjuntos de datos) y la UNED forman el Consorcio Madroño para poner en marcha el repositorio colectivo e-Ciencia-Datos (2). Algo parecido ocurre en Cataluña, donde existe el repositorio CORA.RDR (3). La Universidad de las Islas Baleares también participa en el consorcio catalán (si bien no había publicado conjunto de datos alguno en el momento de la recogida de datos). Así, este segundo consorcio queda formando por ocho universidades.

2.2. Organizaciones

REBIUN es la red que reúne a las bibliotecas universitarias. Es un grupo de trabajo de CRUE (Conferencia de Rectores de las Universidades Españolas). Es la referencia para el desarrollo de los repositorios institucionales de las universidades españolas. Dentro de esta organización existe el grupo de trabajo de repositorios que tiene como objetivo la gestión de los datos de investigación en las universidades españolas. Este grupo es autor de una memoria de buenas prácticas de los servicios ofrecidos (2018) con el objetivo de presentar posibles soluciones para un mismo servicio, ofreciendo modelos a seguir por otras instituciones, recogiendo además ejemplos de iniciativas llevadas a cabo.

Casi todas estas buenas prácticas tienen que ver con las políticas y estrategias seguidas para concienciar a los investigadores de la importancia y necesidad de elaborar proyectos de gestión de datos y considerar casos de buenos ejemplos a seguir. En cuanto a los aspectos tecnológicos, este grupo ha trabajado en un esquema común de metadatos basado en Dublin Core (2017) verificando que “no existen muchos ejemplos de catálogo específico de datos institucional; algunas universidades han resuelto el depósito de datos en su propio repositorio institucional”. Esto, como veremos en la discusión de resultados, es la causa de que casi todas las universidades hayan implementado sus repositorios de datos de investigación con la plataforma DSpace (4).

FECYT es la agencia del ministerio de Ciencia e Innovación que sirve de apoyo a universidades, centros de investigación y otras instituciones para el desarrollo del i+d+i. Su misión principal es ser catalizadora de la relación entre la ciencia y

la sociedad, impulsando el crecimiento de la cultura científica española y fomentando la transferencia de conocimientos. En su organigrama, cuenta con la Unidad de Acceso Abierto que aboga por la eliminación de las barreras que impiden el acceso a los resultados de la investigación científica, mayoritariamente financiada con fondos públicos. Esta unidad es uno de los pilares sobre los que asienta la transición hacia la Ciencia Abierta (UNESCO, 2021) en España y es la redactora de una guía para la evaluación de los repositorios institucionales de investigación que recoge un conjunto de ítems para medir la calidad de los metadatos y un conjunto de aspectos a evaluar de su interoperabilidad (FECYT, 2021). Esta unidad, además, ha tenido un papel decisivo en la elaboración de la Estrategia Nacional de Ciencia Abierta de España (2023).

Dato	Propósito
Tipo	Si se trata de una universidad individual o de un conjunto de universidades parte de un consorcio la que implementa el portal de datos de investigación.
URL consulta datasets	Dirección del portal en la que se pueden consultar los conjuntos de datos depositados.
URL informativa	Dirección donde se encuentra ayuda y documentación sobre el portal de datos de investigación.
Modalidad	Si los datos de investigación se depositan en un portal específicamente implementado para ello o si se depositan dentro del repositorio institucional.
Software	La aplicación informática con la que se ha implementado el portal de datos de investigación.
#datasets	El número de conjuntos de datos depositados.
Archivo	Si el depósito de los conjuntos de datos lo hacen directamente los investigadores o se delega en técnicos de la universidad.
Notas	Documentos de ayuda y guía al usuario. Tutoriales. Infografías. Acceso a herramientas informáticas para la gestión de los datos de investigación.
Fecha/as de recogida de datos	Se han recopilado los datos, en una primera revisión del 6 al 12 de marzo de 2023. Posteriormente, se han revisado datos y anotaciones hasta el día 25 de marzo de 2023.

Tabla I. Recopilación preliminar de información sobre el total de universidades públicas españolas

2.3. Recogida inicial de información

En una primera instancia se ha trabajado sobre las 49 universidades indicadas anteriormente, las que forman un consorcio (15 en total) y las que han dispuesto repositorios de datos de investigación de forma individual (34), tomándose los datos que se muestran en la Tabla I (a la izquierda).

La revisión inicial de los repositorios de conjuntos de datos de investigación muestra, en general, un nivel de implementación muy diverso, oscilando desde instalaciones mínimas hasta la implementación de repositorios colectivos de consorcios de instituciones de educación superior. El número de universidades con menos de una decena de conjuntos de datos depositados asciende a 16, lo que lleva a pensar, con cierta base, que la gestión para el depósito de estos conjuntos de datos se encuentra aún en una fase embrionaria. Este conjunto lo forman: Cantabria, Córdoba, Huelva, Jaén León, Oviedo, Pablo de Olavide, Politécnica de Cartagena, Castilla-La Mancha, Extremadura, La Rioja, Miguel Hernández, Pública de Navarra, Santiago y Vigo. Las siete últimas no disponían de ningún conjunto de datos publicado cuando se llevó a cabo el análisis de los repositorios, así que se excluyeron del estudio y se redujo el total de universidades analizadas de forma individual a 25.

2.4. Determinación de la muestra de estudio

Se han tomado las siguientes decisiones para formar la muestra objeto de estudio: se analizan los consorcios como si de una única institución se tratara, se excluyen del estudio aquellas universidades con cero conjuntos de datos depositados, y se excluyen del estudio a aquellas universidades que no implementen su repositorio de datos de investigación con software libre. De esta forma, la muestra queda determinada por el conjunto de universidades recogido en la *Tabla II*.

Nombre	#univ	Composición	#datasets
CORA	8	Autónoma de Barcelona; Barcelona; Girona; Lleida; Oberta de Catalunya; Politécnica de Catalunya; Pompeu Fabra; Rovira i Virgili e Illes Balears	467
e-Ciencia-Datos	6	Alcalá de Henares; Autónoma de Madrid; Carlos III, UNED; Politécnica de Madrid y Rey Juan Carlos	744
Resto	25	A Coruña; Alicante; Almería; Burgos, Cádiz, Cantabria, Complutense, Córdoba, Granada, Huelva, Jaén, Jaime I, León, Málaga, Murcia, Oviedo, Pablo de Olavide, País Vasco, Politécnica de Cartagena, Politécnica de Valencia, Salamanca, Sevilla, Valencia, Valladolid y Zaragoza	744

Tabla II. Composición de la muestra objeto de estudio y número de datasets

3. Discusión de resultados

A partir de los aspectos analizados en la revisión inicial de los repositorios de datos de investigación de las universidades que forman parte de la muestra, se obtienen los siguientes resultados:

3.1. Tipo de organización

Son dos los consorcios y 24 las universidades que han implementado repositorios de datos de investigación con, al menos, un conjunto de datos publicado. Los consorcios agrupan 14 universidades con conjuntos de datos publicados.

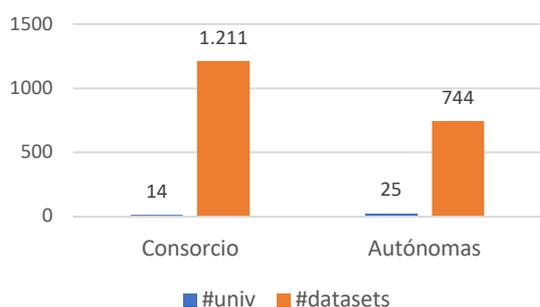


Figura 1. Total de repositorios y conjuntos de datos por tipo de organización (consorcio/universidad)

Las diez universidades que más conjuntos de datos tienen publicados, en sus repositorios institucionales o en uno colectivo, son las siguientes:

Universidad	#datasets
Alcalá de Henares	375
Carlos III	259
Zaragoza	222
Granada	112
Barcelona	105
Politécnica de Catalunya	75
Lleida	59
Pompeu Fabra	44
Politécnica de Valencia	43
Politécnica de Madrid	41

Tabla III. Las 10 primeras universidades españolas por número de conjuntos de datos publicados

Estas diez universidades totalizan 1.334 conjuntos de datos, el 68,3% del total de conjuntos publicados por las universidades públicas españolas (5). Las otras 28 universidades analizadas publican un poco más del 30%. Existe, pues, una

fuerte concentración en las universidades de la tabla anterior, mucho más manifiesta entre las que superan los 100 conjuntos publicados (cinco). Destaca el hecho de que las tres universidades politécnicas aparecen en la Tabla III. También es interesante el dato de que siete de estas universidades formen parte de consorcio (tres del madrileño y cuatro del catalán).

3.2. Año de publicación

La distribución de conjuntos de datos publicados año por año se ve en el siguiente gráfico.

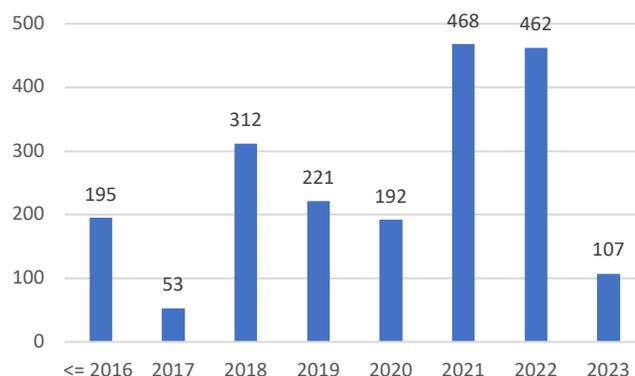


Figura 2. Conjuntos de datos por año de publicación

Los años 2021 y 2022 son los que se han publicado más conjuntos de datos en las universidades españolas, si bien ese dato está muy influenciado porque la Universidad de Zaragoza solo ha publicado en esos dos años (81 y 141 conjuntos de datos, todos ellos sobre la calidad del aire, lo que también influye en el análisis de las materias). Se ha incluido el total de conjuntos de datos de 2023 que, obviamente será mucho mayor cuando finalice el año en curso ya que muestra una proyección, de seguir el ritmo actual, de una cifra similar a la de los dos últimos años

Llama la atención el alto número de conjuntos de datos publicados en el año 2016 o antes. Esto se debe a un caso particular, el de la Universidad de Sevilla, en la que dieron de alta ese año un total de 124 conjuntos de datos de arquitectura y urbanismo del Protectorado Español en Marruecos, fondos documentales donados por el Ejército Español. El conjunto de datos se crea en esa fecha, si bien la información que contienen es de fechas muy anteriores (entre los años 1930 y 1950). En todo caso, queda pendiente confirmar. También destacan los conjuntos de datos de investigaciones bibliométricas de la Universidad de Granada, la mayor parte de ellas realizadas antes de ese año. Es previsible una tendencia positiva en la publicación de conjuntos de datos este año.

3.3. Materia

Se han revisado los conjuntos de datos publicados por las universidades y consorcios de la muestra y se han asignado a las cinco ramas tradicionales de investigación en el sistema universitario español: artes, salud, ciencias, sociales e ingenierías. Los resultados se visualizan en el siguiente gráfico:

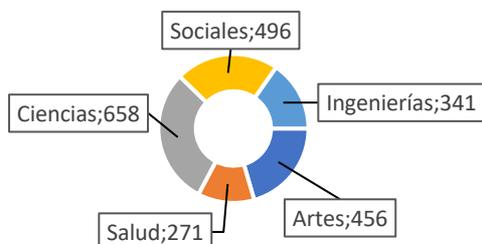


Figura 3. Conjuntos de datos por año de publicación

El 30% de los conjuntos de datos publicados son de ciencias, seguidos de ciencias sociales y artes y humanidades. En ciencias se nota la influencia de la Universidad de Zaragoza donde todos los conjuntos de datos son de la misma temática (medio ambiente) y están publicados solo en dos años. En artes y humanidades casi el 75% lo aporta el consorcio Madroño (Universidad de Alcalá de Henares). Recordemos también el caso de la Universidad de Sevilla y los datos de arquitectura y urbanismo de ciudades norteafricanas, casi un tercio del total. En ciencias sociales destaca la aportación del consorcio madrileño (269 conjuntos de datos) y, en particular, de la Universidad de Granada con un alto número de conjuntos de datos bibliométricos publicados por el grupo de investigación EC3 (la mayoría antes del año 2016).

3.4. Plataforma software

Los dos consorcios han implementado sus repositorios usando la plataforma Dataverse. Por regla general, las universidades que han acometido el desarrollo del repositorio de datos de investigación por su cuenta emplean el software DSpace. Una de ellas, las Palmas de Gran Canaria, emplea DSpace-CRIS, la distribución de DSpace específica para la gestión de datos de investigación (aunque solo se habían depositado dos conjuntos en el momento del estudio). Cuando se usa DSpace, el repositorio de los conjuntos de datos suele ser una colección específica del repositorio institucional. Algunas universidades están ahora migrando a la versión 7 de este sistema que aporta varias prestaciones para la gestión de conjuntos de datos de investigación.

La Universidad de A Coruña no ha implementado su repositorio, sino que redirige a los investigadores a Zenodo, el repositorio de acceso abierto de propósito general desarrollado por el CERN dentro de una iniciativa de la Comisión Europea. La Universidad de Zaragoza emplea la plataforma Invenio RDM para implementar su repositorio que se da la circunstancia de que este software es una distribución de Zenodo. Por último, la otra universidad canaria, La Laguna, emplea tecnología de pago (Digital Commons de Elsevier), por lo que ha quedado excluida del estudio.

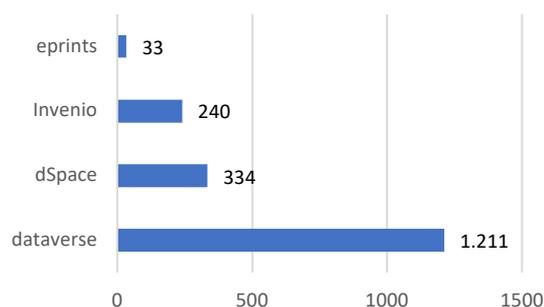


Figura 4. Conjuntos de datos distribuidos según la plataforma software en la que se han implementado

3.5. Archivo delegado o autoarchivo

Revisando las páginas de información de los repositorios de datos de investigación (también las de los repositorios institucionales cuando en las primeras no aparecía esa información), se han identificado tres modalidades: autoarchivo, archivo delegado en el personal del repositorio y ambas modalidades al mismo tiempo. No ha sido sencillo encontrar esta información y en algún caso ha resultado imposible. En el caso de los consorcios el archivo es delegado (es uno de los servicios ofrecidos). En el caso de las universidades que gestionan de forma individual el repositorio, 12 de ellas optan por el autoarchivo, 1 por el archivo delegado y 3 por la combinación de ambos. El predominio del autoarchivo es fruto, sin duda alguna, de la costumbre heredada del depósito en los repositorios institucionales. En algunas universidades se informa a los investigadores de posibles futuras revisiones del depósito para verificar la calidad del mismo.

3.6. Ayuda y orientación

La apuesta decidida de Rebiun y FECYT por la Ciencia Abierta sí parece haber tenido éxito en las bibliotecas universitarias en lo relacionado con la ayuda y orientación a la comunidad investigadora por medio de biblioguías, infografías, información sobre los planes de gestión de datos,

aplicaciones para desarrollarlos, plantillas para hacerlo manualmente y aprendizaje de cómo deben ser citados.

Tras la revisión del total de universidades objeto del estudio, podemos afirmar que 36 de 39 orientan y ayudan de forma suficiente a su comunidad. Es previsible que las tres universidades restantes solucionen este problema pronto. Los temas más presentes en estos contenidos son los siguientes:

- La gestión de los conjuntos de datos de investigación. En qué consiste y cuál es su finalidad.
- Los planes de gestión de datos, qué son y para qué sirven. Guías para la elaboración de estos planes.
- Información sobre aplicaciones para la elaboración de los planes de gestión de datos. Se recomienda DMP Online, ARGOS y PGD Online (aplicación del consorcio madrileño).
- Orientación sobre dónde publicar los conjuntos de datos. Se citan los principales repositorios generalistas, en especial Zenodo y el repositorio institucional de cada universidad (con algo menos de frecuencia, parece darse por sentado que los investigadores saben que disponen de ese servicio).
- En algunas universidades se informa de la importancia del uso de licencias de libre acceso a los contenidos (Creative Commons generalmente).

4. Análisis y descripción de las plataformas software empleadas

En la revisión de los repositorios de datos de investigación en funcionamiento se han identificado dos plataformas software ampliamente utilizadas (Dataverse y DSpace) y otra tres de menor frecuencia de uso (que en realidad son dos porque Invenio RDM y/o Zenodo son distintas implementaciones del mismo software).

4.1. Dataverse

Es una aplicación web de código abierto para compartir, preservar, citar, explorar y analizar datos de investigación. Facilita poner los datos a disposición de otros y permite replicar el trabajo de otros más fácilmente. Este proyecto está liderado por el Instituto de Ciencias Sociales Cuantitativas de la Universidad de Harvard y se basa en la experiencia de un anterior proyecto del Centro Virtual de Datos que se desarrolló entre 1997 y 2006 como una colaboración entre centro de datos Harvard-MIT (ahora parte de este instituto) y la Biblio-

teca de la Universidad de Harvard. Cada colección contiene conjuntos de datos, y cada conjunto de datos contiene metadatos descriptivos y archivos de datos (incluida la documentación y el código que acompañan a los datos). Como método de organización, estas colecciones también pueden contener otras colecciones (6).

4.2. DSpace

Se trata de una aplicación de repositorio web que permite a investigadores y académicos publicar documentos y datos. Aunque comparte algunas características con los sistemas de gestión documental, este software responde a una necesidad específica como sistema de archivos digitales, centrado en el almacenamiento a largo plazo, el acceso y la preservación digital. Este sistema ha tenido muy buena acogida en el seno de las organizaciones académicas para crear sus proyectos de repositorio digital de acceso abierto. Es gratuito, fácil de instalar y personalizable. Un punto fuerte es su amplia comunidad de desarrolladores en constante crecimiento, comprometida con la ampliación y mejora continuas.

La primera versión pública de DSpace se publicó en noviembre de 2002, en un esfuerzo conjunto de desarrolladores del MIT y HP Labs que terminaron poniendo en marcha de forma conjunta la DSpace Foundation. Actualmente el trabajo de la comunidad y el mantenimiento del software está liderado por LYRISIS, entidad que procede de la antigua fundación (7).

4.3. eprints

Es una plataforma de repositorios digitales de código abierto. Desarrollada en la Universidad de Southampton en el año 2000. Es software gratuito y de código abierto para crear repositorios de acceso abierto que cumplan el protocolo OAI-PMH. Comparte muchas características con los sistemas de gestión de documentos, pero su uso principal es de repositorio institucional y portal de revistas científicas (8).

4.4. InvenioRDM y Zenodo

Éste es un caso particular porque, en esencia, es el mismo software con distintas distribuciones. El repositorio de datos de investigación Zenodo, desarrollado por el CERN dentro del programa OpenAIRE (9), se ejecuta básicamente bajo la versión 3 de Invenio “envuelta” por una pequeña capa extra de código (que también se llama Zenodo). Para simplificar la reutilización de la base de código Zenodo, varias instituciones se unieron en 2019 para distribuir un paquete agnóstico para instituciones bajo el nombre de InvenioRDM, un marco de desarrollo de código

abierto para repositorios digitales a gran escala centrado específicamente en la seguridad y en la conservación a largo plazo (10).

Subgrupo FAIR	Significado
F: Localizable	C1 ¿Se puede acceder a un determinado conjunto de datos de investigación en una versión actual a través de un PID/DOI único?
	C2 ¿Está la información de los datos de investigación depositados indexada en catálogos de datos, registros y motores de búsqueda?
	C3 ¿Existe una interfaz de búsqueda con posibilidades de filtrado para datos enlazados estructurados?
A: Accesible	C4 ¿Pueden almacenarse o referenciarse fácilmente nuevos datos de investigación?
	C5 ¿Es la interfaz de entrada de los datos fácil de usar ocultando términos técnicos e identificadores?
	C6 ¿Se puede acceder directamente a los datos y/o metadatos de la investigación vía http(s)?
	C7 ¿Existen parámetros de autenticación y autorización para el acceso de los usuarios?
I: Interoperable	C8 ¿Está disponible la descripción de los metadatos en una serialización RDF?
	C9 ¿Pueden utilizarse determinadas ontologías establecidas para describir el conjunto de datos de investigación de forma general (schema.org/Dataset, DataCite o DCAT/DublinCore)?
	C10 ¿Pueden utilizarse vocabularios específicos de un dominio para describir con más detalle los datos de investigación?
	C11 ¿Puede describirse cada concepto relacionado con el conjunto de datos de investigación con su correspondiente URI?
R: Reutilizable	C12 ¿Puede especificarse una licencia de datos en forma de Linked Data?
	C13 ¿Puede especificarse y actualizarse la procedencia de los datos de forma estructurada?
	C14 ¿Se relacionan los conjuntos de datos basándose en Linked Data y en criterios como tema, comunidad, métodos utilizados o similares?
	C15 ¿Se validan los datos proporcionados o existen controles de conformidad?

Tabla IV. Criterios de evaluación de las plataformas de gestión de conjuntos de datos basados en los principios FAIR (Langer et al., 2019)

Langer et al. (2019) analizaron diversas plataformas para implementar repositorios. En su análisis establecieron tres categorías: *genéricas* para una gestión básica de los datos; *aplicaciones* como desarrollo de estas plataformas y *otras herramientas* para la gestión de datos. Las cuatro plataformas que se usan en las universidades españolas pertenecen a la primera categoría. Posteriormente investigaron sus capacidades para la aplicación de una gestión de datos interdisciplinaria sostenible. Para comparar las plataformas establecieron 15 criterios de evaluación a partir de los principios FAIR para la gestión de datos de investigación (Alcalá y Anglada, 2019), que se centran particularmente en el conocimiento y manejo de los enlaces de datos, más en concreto en el conocimiento y la gestión de los datos abiertos enlazados en estas soluciones de publicación de datos. La siguiente tabla recoge los criterios establecidos por Langer et al. (2019) en su análisis:

Langer et al. (2019), en su análisis, definieron seis posibles resultados para la evaluación del cumplimiento de cada criterio.

Indicador	Significado
+	El criterio se cumple íntegramente
O	El criterio se cumple parcialmente
-	El criterio no se cumple
%	El criterio no es aplicable
¿?	No fue posible evaluar ese criterio
()	La característica está parcialmente desarrollada en la versión original de la plataforma, pero puede ampliarse desarrollando código

Tabla V. Criterios de evaluación del cumplimiento de los principios FAIR por el software de repositorios digitales (Langer et al., 2019)

Plataforma	F	A	I	R
Dataverse	C1+	C4+	C8+	C12+
	C2+	C5+	C9+	C13+
	C3o	C6+	C11o	C14o
		C7o		
DSpace	C1+	C4o	C9+	C12o
	C2+	C5o		C13+
	C3o	C6+		C14o
		C7+		
eprints	C1+	C4+		C12+
	C2+	C6+		C13+
	C3o	C7o		C14o
Invenio	C1 +	C4+	C9o	C13o
	C2 +	C6+	C11o	C15+
	C3 o			

Tabla VI. Cumplimiento de los principios FAIR por parte de las cuatro plataformas de repositorio digital analizadas (Langer et al., 2019)

En la tabla anterior se recoge un resumen parcial de los resultados de cada plataforma. A efectos de este estudio, únicamente interesan los resultados positivos (“+”), parcialmente positivo (“o”) y potencialmente positivo (“()”), porque representan el cumplimiento de algún criterio FAIR. DataVerse es la plataforma que más principios FAIR satisface (9 de forma íntegra y 4 parcialmente) y se puede mejorar en los criterios C11 y C14 desarrollando software adicional. El único criterio que no se cumple, C10, tiene que ver con el posible uso de un vocabulario específico de un dominio para la descripción de los conjuntos de datos, criterio que no cumple ninguna de las cuatro plataformas en realidad. En el caso de las otras plataformas no se aprecian diferencias significativas entre ellas.

5. Conclusiones

En cuanto a la implementación de los repositorios de conjuntos de datos de investigación en las universidades públicas españolas, estamos prácticamente al principio de su desarrollo. Algunas prácticamente no han empezado. El escaso número de conjuntos de datos publicados desvirtúa el análisis cuantitativo. Hay universidades donde solo se han publicado dos o tres conjuntos de datos. Los consorcios ayudan a visibilizar estas fuentes de información. Las comunidades autónomas con varias universidades (Andalucía, Castilla y León y la Comunidad Valenciana principalmente), deberían plantearse esta opción como una clara oportunidad de mejora.

El estado embrionario de esta implantación es una oportunidad para unificar criterios que eviten la multiplicidad de esfuerzos similares y la dispersión de los desarrollos e implementaciones, algo que ya ha ocurrido con los repositorios generalistas y que, difícilmente, tiene vuelta atrás.

El software de repositorio digital que mejor cumple con los principios FAIR es DataVerse. Es el que usan los consorcios, con lo cual, un porcentaje importante de universidades satisfacen esos principios de apertura e interoperabilidad. DSpace es el más instalado, aunque el total de conjuntos de datos gestionados por esta plataforma es menor que el gestionado por los consorcios (otra razón más para la búsqueda de la gestión colaborativa).

En casi todos los casos existen guías de orientación y ayuda a los investigadores sobre los planes generales de datos y cómo depositarlos por medio de “bibliografías” dentro de la web de la biblioteca universitaria.

El depósito de los conjuntos de datos se lleva a cabo, de forma mayoritaria, por medio del autoarchivo, revisado (en algunos casos) por el personal técnico de la biblioteca. Quizá esa revisión debería ser obligatoria para mejorar la descripción.

Resulta innegable que una gestión de los conjuntos de datos, acompañada de una campaña de sensibilización a favor de su uso entre la comunidad investigadora, reforzará el papel de las bibliotecas universitarias como servicio de apoyo a la investigación. Un problema típico es la posible confusión que puede tener un investigador que haya recibido financiación europea para su investigación. En principio, sus conjuntos de datos deberían ser depositados en Zenodo, con independencia de publicar una copia en su universidad. Lo cierto es que se debe reflexionar sobre si es conveniente y necesaria esa duplicación de espacio de almacenamiento (algo poco sostenible que puede terminar siendo un problema), si bien puede ser interesante a efectos de la recuperación de información. En España, el marco legal definido por la Ley de la Ciencia (2022) también impone la publicación, así que perfilar políticas claras y sostenibles en el tiempo es algo sobre lo que se debe comenzar a trabajar pronto.

Los repositorios de conjuntos de datos de investigación deberían ofrecer métricas en abierto que permitan establecer si se está cumpliendo el objetivo de la reutilización de los datos. Es una de las líneas de investigación más necesarias porque va a permitir medir si todos los esfuerzos que se están desarrollando en este momento ayudan al desarrollo de nuevas investigaciones (Barrett et al, 2021).

Notas

- (1) CC son las siglas de las licencias ‘Creative Commons’ que tienen varias especificaciones: CC-BY permite copiar y redistribuir el material en cualquier medio o formato, remezclando y transformando el contenido para cualquier propósito, incluso comercialmente. CC0 es bastante parecida, se renuncia a todos los derechos a la obra bajo las leyes de derechos autorales en todo el mundo. Más información sobre estas licencias en <https://creativecommons.org>
- (2) eCienciaDatos es el repositorio de datos de investigación del Consorcio Madroño. Más información en <https://datos.consorciojadrono.es/>
- (3) CORA.RDR es el repositorio de datos federado y multidisciplinar para la publicación de conjuntos de datos de investigación en modo FAIR, siguiendo las directrices de la European Open Science Cloud (EOSC). Más información en <https://www.csuc.cat/es/servicios/cora-repositorio-de-datos-de-investigacion>
- (4) DSpace es un software de código abierto que provee herramientas para la administración de colecciones digitales, y comúnmente es usada como solución de repositorio bibliográfico institucional. Soporta una gran variedad de

datos, incluyendo libros, tesis, fotografías, filmes, video, datos de investigación y otras formas de contenido. Los datos son organizados como ítems que pertenecen a una colección; cada colección pertenece a una comunidad. Fue liberado bajo una licencia BSD en el 2002, como producto de una alianza de Hewlett Packard y el MIT.

- (5) En el estudio se ha analizado el número de conjuntos de datos presentes en los repositorios institucionales. Puede ser que una universidad disponga de una colección en Zenodo o en otro repositorio, como es el caso de A Coruña, si bien solo hemos identificado esta universidad. El total de documentos no ha de variar mucho.
- (6) La página del proyecto Dataverse está en la URL: <https://Dataverse.org/>
- (7) Más información sobre DSpace en la URL: <https://dspace.lyrasis.org/>
- (8) Más información sobre este proyecto en <https://www.eprints.org/uk/>
- (9) OpenAIRE es una infraestructura tecnológica y de servicios para apoyar, acelerar y medir la correcta implementación de las políticas europeas de acceso abierto a publicaciones científicas y datos de investigación. Cuenta con una sólida red de agentes nacionales que actúan como puntos de referencia nacional para divulgar y difundir las políticas hacia la Ciencia Abierta auspiciadas por la Comisión Europea entre las instituciones y los investigadores, así como para facilitar la coordinación de las políticas nacionales con las europeas. El agente español es FECYT. Más información está disponible en <https://recolecta.fecyt.es/open-aire>
- (10) La página web del paquete de código abierto Invenio-RDM es <https://inveniosoftware.org/>

Referencias

- Abadal, E.; et al. (2023). Ciencia abierta en España 2023: informe de situación y análisis de la percepción. <http://hdl.handle.net/2445/200020>
- Alcalá, M.; Anglada, L. (2019). FAIR x FAIR: Requisitos factibles, alcanzables e implementables para un repositorio de datos de investigación FAIR. https://www.recercat.cat/bitstream/handle/2072/356460/InformeFxF_maquetada_ESP.pdf
- Alonso-Arévalo, J. (2019). La gestión de datos de investigación en el horizonte de las bibliotecas universitarias y de investigación. // Cuadernos de Documentación Multimedia. 30, 75-88. <https://doi.org/10.5209/CDMU.62806>
- Angelozzi, S. M. (2020). La gestión de datos de investigación en abierto: introducción al rol emergente para las bibliotecas universitarias y científicas argentinas. // Palabra clave. 9:2, e091. <https://doi.org/10.24215/18539912e091>
- Ashiq, M.; Usmani, M. H.; Naeem, M. (2022). A systematic literature review on research data management practices and services. // Global Knowledge, Memory and Communication. 71:8/9, 649-671. <https://doi.org/10.1108/GKMC-07-2020-0103>
- Ayris, P.; Ignat, T. (2018). Defining the role of libraries in the Open Science landscape: a reflection on current European practice. // Open Information Science. 2:1, 1-22. <https://doi.org/10.1515/opis-2018-0001>
- Barrett; et al. (2021) Metrics for Data Repositories and Knowledge bases: Working Group Report. <https://datascience.nih.gov/sites/default/files/Metrics-Report-2021-Sep15-508.pdf>
- Bethencourt-Aguilar, A.; Castellanos-Nieves, D.; Sosa-Alonso, J. J.; Area-Moreira, M. (2022). Implicaciones técnicas y prácticas de las Redes Adversarias Generativas a la Ciencia Abierta en Educación. // RiITE Revista Interuniversitaria de Investigación en Tecnología Educativa, 138-156. <https://doi.org/10.6018/riite.545881>
- Borghini, J.A.; Van Gulick, A.E. (2021). Promoting Open Science through research data management. // arXiv preprint <https://arxiv.org/abs/2110.00888>
- De Giusti, M. R. (2021) Calidad en los repositorios digitales: los principios TRUST para repositorios de datos. // Revista Iberoamericana de Tecnología en Educación y Educación en Tecnología. 29, 55-59 <https://doi.org/10.24215/18509959.29.e6>
- España (2022). Ley 17/2022, de 5 de septiembre, por la que se modifica la Ley 14/2011, de 1 de junio, de la Ciencia, la Tecnología y la Innovación. // Boletín Oficial del Estado (2022-09-05). <https://www.boe.es/eli/es/l/2022/09/05/17/con>
- España (2023). Estrategia Nacional de Ciencia Abierta (EN-CA). <https://www.ciencia.gob.es/InfoGeneralPortal/documento/c30b29d7-abac-4b31-9156-809927b5ee49>
- European Commission (2014). Guía del participante Horizonte 2020. <https://www.horizonteeuropa.es/sites/default/files/inline-files/guia-del-participante-h2020.pdf>
- European Commission (2022). Guía del participante Horizonte Europa. https://www.horizonteeuropa.es/sites/default/files/noticias/Gu%C3%ADa%20del%20participante%20-%20Horizonte%20Europa%20web_0.pdf
- European Commission (2023). Políticas de acceso abierto en América Latina, el Caribe y la Unión Europea: avances para un diálogo político. <https://data.europa.eu/doi/10.2777/162>
- FAIR Data Maturity Model Working Group (2020). FAIR Data Maturity Model. Specification and Guidelines (1.0). <https://doi.org/10.15497/rda00050>
- FECYT (2021) Guía para la evaluación de repositorios institucionales de Investigación. <https://www.fecyt.es/es/publicacion/guia-para-la-evaluacion-de-repositorios-institucionales-de-investigacion>
- Federer, L. M.; Qin, J. (2019). Beyond the data management plan: Expanding roles for librarians in data science and open science. // Proceedings of the Association for Information Science and Technology. 56:1, 529-531. <https://asistdl.onlinelibrary.wiley.com/doi/abs/10.1002/pr.a2.82>
- Fundación CYD. (2023). La investigación en España: aumenta el gasto y la producción científica, pero con recursos inferiores a países de su entorno. Informe CYD. <https://www.fundacioncyd.org/la-investigacion-en-espana-aumenta-el-gasto-y-la-produccion-cientifica-pero-con-recursos-inferiores-a-paises-de-su-entorno/>
- Johnston, L.; Carlson, J.; Hswe, P.; Hudson-Vitale, C.; Imker, H.; Kozlowski, W.; Olendorf, R. and Stewart, C. (2017). Data Curation Network: How Do We Compare? A Snapshot of Six Academic Library Institutions' Data Repository and Curation Services. // Journal of eScience Librarianship. 6:1. e1102. <https://doi.org/10.7191/jeslib.2017.1102>
- Johnston, L.; Carlson, J.; Hswe, P.; Hudson-Vitale, C.; Imker, H.; Kozlowski, W.; Olendorf, R.; Stewart, C.; Blake, M.; Herndon, J.; McGear, T. and Hull, E. (2018). Data Curation Network: A Cross-Institutional Staffing Model for Curating Research Data. // International Journal of Digital Curation. 13:1. <http://www.ijdc.net/article/view/616>
- Langer, A.; Bilz, E.; Gaedke, M. (2019). Analysis of Current RDM Applications for the Interdisciplinary Publication of Research Data. <https://ceur-ws.org/Vol-2447/paper1.pdf>
- López Carreño, R.; Martínez Méndez, F. J. (2020). Sistemas de recuperación de información implementados a partir de CORD-19: herramientas clave en la gestión de la in-

- formación sobre COVID-19. // Revista Española de Documentación Científica. 43:4. <https://redc.revistas.csic.es/index.php/redc/article/view/1300>
- Marín-Araiza, P.; Puerta-Díaz, M. y Gregorio-Vidotti, S. (2019). Gestión de datos de investigación y bibliotecas: preservando los nuevos bienes científicos. // *Hipertext.net*. 19, 13-31. <https://raco.cat/index.php/Hipertext/article/view/360098>
- OECD/LEGAL/0347 (2021). Recommendation of the Council concerning Access to Research Data from Public Funding. <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0347>
- REBIUN (2017) Plantilla de metadatos para la descripción de datos de investigación. // Estudio de los repositorios de datos, 2017. <http://hdl.handle.net/20.500.11967/137>
- REBIUN (2018) Gestión de datos de investigación en las universidades españolas y en el CSIC: memoria de buenas prácticas de los servicios ofrecidos. https://www.rebiun.org/sites/default/files/Gestion_Datos_Memoria_buenas_Practicas_201811.pdf
- Rocca-Serra, P.; et al. (2023). The FAIR Cookbook: the essential resource for and by FAIR doers. // *Scientific Data*. 10, 292. <https://doi.org/10.1038/s41597-023-02166-3>
- Sheikh, A.; Malik, A.; Adnan, R. (2023). Evolution of research data management in academic libraries: A review of the literature. // *Information Development*. 0:0. <https://doi.org/10.1177/02666669231157405>
- Redkina, N. S. (2019). Current Trends in Research Data Management. // *Scientific & Technical Information Processing*. 46:2, 53-58. <https://doi.org/10.3103/S0147688219020035>
- Torres-Salinas, D.; Robinson-García, N.; Castillo-Valdivieso, P. A. (2020). Open Access and Altmetrics in the pandemic age: Forecast analysis on COVID-19 literature. // *BioRxiv*. 2020-04. <https://doi.org/10.1101/2020.04.23.057307>
- UNESCO (2021) Recomendación de la UNESCO sobre la Ciencia Abierta. https://unesdoc.unesco.org/ark:/48223/pf0000379949_spa (2023-03-02)
- Wilkinson, M. D.; Dumontier, M.; Aalbersberg, Ij. J.; Appleton, G.; Axton, M.; Baak, A.; y Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. // *Scientific Data*. 3, 160018. <https://doi.org/10.1038/sdata.2016.18>

Enviado: 2021-04-01. Segunda versión: 2021-07-25.
Aceptado: 2021-11-02.

Análisis reticular de las relaciones de colaboración de las empresas de desechos tecnológicos de Panamá

Network analysis of collaborative relationships of WEEE companies in Panama

Ángel ZAZO (1), Saúl ARDINES (2), Amarilis DE LEÓN (2), María Luisa PAZ (2),
Virginia JUÁREZ (2), Isaac PERDOMO (2)

(1) Universidad de Salamanca, Instituto Universitario de Estudios de la Ciencia y de la Tecnología, C/ Espejo, 2, 37008, Salamanca (España), angelzazo@usal.es. (2) Universidad de Panamá. Campus Dr. Octavio Méndez Pereira, Vía Transistmica, Ciudad de Panamá (Panamá). {saul.ardines | amarilis.deleon | maria.pazf | virginia.juarez | isaac.perdomo }@up.ac.pa

Resumen

Este artículo describe los resultados de una investigación exploratoria llevada a cabo para identificar las relaciones que las empresas que se dedican a la recolección y/o reciclado de residuos de aparatos eléctricos y electrónicos (RAEE) en la República de Panamá mantienen con diferentes entidades del país y extranjeras. A partir de las respuestas dadas a un cuestionario se han aplicado dos enfoques distintos: el análisis de redes sociales (ARS) y el análisis reticular de coincidencias (ARC). Las pocas relaciones obtenidas aplicando el ARS nos muestran que la transferencia de conocimiento, información y recursos entre entidades, que sería deseable, es, lamentablemente, muy escasa. Con el ARC hemos analizado cuáles son las características de las relaciones que han mantenido las empresas de desechos tecnológicos, su relación con el tipo de entidad con la colaboran, la pertenencia a grupos de empresas nacionales o extranjeros y cómo se gestiona esa colaboración. Se ha evidenciado que existe una escasa concienciación sobre la recolección y reciclado de los desechos tecnológicos y que es necesaria la implicación gubernamental en el desarrollo de leyes y normativas sobre los RAEE, que regulen los procesos técnicos y las vinculaciones entre todos los actores que intervienen en el proceso, así como medidas que faciliten la colaboración entre las empresas, tanto en los aspectos administrativos y regulatorios como mediante incentivos adecuados para su impulso.

Palabras clave: Desechos tecnológicos. RAEE. Panamá. Análisis de redes sociales. ARS. Análisis reticular de coincidencias. ARC.

1. Introducción

Los residuos de aparatos eléctricos y electrónicos (RAEE) se consideran actualmente uno de los tipos de residuos de más rápido crecimiento en el mundo, con una tasa de crecimiento estimada que va del 3 % al 5 % anual (Forti et al., 2020; Shittu et al., 2021). Son innumerables los trabajos que indican los problemas medioambientales y para la salud que supone la desidia en el tratamiento adecuado de los RAEE (Prasad

Abstract

In this article we describe the results of an exploratory research carried out to learn about the relationships that companies involved in the collection and/or recycling of waste electrical and electronic equipment (WEEE) in the Republic of Panama maintain with different entities in the country and abroad. Based on the answers given to a questionnaire, two different approaches have been applied: social network analysis (SNA) and network analysis of coincidences (NAC). The few relationships obtained by applying SNA show us that the transfer of knowledge, information, and resources between entities, which would be desirable, is unfortunately very rare. With the NAC we have analysed the characteristics of the relationships that the WEEE companies have maintained, their relationship with the type of entity with which they collaborate, whether they belong to national or foreign groups of companies and how this collaboration is managed. It has become evident that there is little awareness of collection and recycling of technological waste and that there is a need for government involvement in the development of laws and regulations on WEEE, which regulate the technical processes and the relationships between all the actors involved in the process, as well as actions to facilitate collaboration between companies, both in administrative and regulatory aspects and through appropriate incentives for their promotion.

Keywords: E-Waste. WEEE. Panamá. Social network analysis. SNA. Network analysis of coincidences. NAC.

et al., 2020). El reciclado de estos desechos permitiría disminuir el uso de recursos vírgenes en la fabricación y, en consecuencia, contribuir a reducir la contaminación ambiental.

Según varios informes internacionales (Forti et al., 2020; Naciones Unidas, 2022; Wagner et al., 2022), en América Latina y El Caribe, más del 97% de los residuos electrónicos generados no se recogen ni se envían a instalaciones espe-

cíficas para una adecuada gestión ambientalmente racional. Las infraestructuras necesarias en países de ingresos bajos y medios todavía no se han desarrollado o son insuficientes para gestionar estos residuos. La mayor parte de los RAEE acaban en vertederos o son gestionados por el sector informal, que separa las partes valiosas que contienen los residuos electrónicos, mientras que los plásticos con COP (contaminantes orgánicos persistentes) se eliminan en vertederos sin ningún tratamiento, con el inminente peligro medioambiental y para la salud humana.

Aunque hay avances legislativos, la realidad es que éstos son muy lentos, la aplicación es muy desigual y, en general, no se estimula la recolección y el manejo adecuado de los desechos electrónicos debido a la falta de inversión y motivación política, que hacen que sean gestionados de manera informal e insegura.

Los desechos tecnológicos proceden del descarte de un aparato eléctrico y/o electrónico. En general, los consumidores tienen una vaga idea de que dichos descartes puedan ser reutilizados y su potencial económico, pues contienen metales preciosos como oro, plata, cobre, platino y paladio, además de otros materiales como hierro, aluminio y plástico que pueden ser reciclados y reutilizados (Baldé et al., 2017).

La situación actual se ha visto agravada debido al virus SARS-COVID-19, dado que el uso de la tecnología se vio incrementada en todos los ámbitos, salud, educación, comunicación, entretenimiento, comercio, etc., lo que ha provocado que la cantidad de desechos tecnológicos generados haya aumentado sustancialmente desde 2020. A ello sumamos el hecho de que cada vez es mayor el consumo de aparatos eléctricos y electrónicos que, además, poseen ciclos de vida muy cortos y diseños que no permiten su reparación o reutilización.

En términos generales la situación de Panamá es similar a lo que sucede en países de su entorno: grandes volúmenes de desechos tecnológicos versus pocas legislaciones y pocas regulaciones operativas al respecto. Según el informe de (Wagner et al., 2022) para América Latina, los RAEE generados en Panamá en 2019 fueron de unas 36.000 toneladas, de las que casi la tercera parte fueron plásticos, de las cuales oficialmente solo se recogieron 80 toneladas. En ese informe se evidencia también que Panamá no cuenta con una herramienta legislativa específica relativa a los RAEE y los COP de estos.

Panamá hizo un amplio estudio de la situación en relación con el Plan Nacional de Gestión Integral de Residuos (PNGIR) 2017-2027 (AAUD & INECO, 2017), que mostró que la gestión de los

residuos estaba considerada como el segundo problema del país, tan solo por detrás de la delincuencia o la inseguridad ciudadana, y que más del 58 % de la población no clasificaba o separaba los residuos reciclables que generaba en su hogar. Se vio que la comunidad no había sido preparada y orientada en ello y, tampoco, en el sistema de recolección. Por ende, no se contaba con personal calificado, ni herramientas y ni áreas para el manejo adecuado de los residuos. Hoy en día pocos cambios de aprecian, y la situación de los RAEE no es diferente.

Pese al PNGIR, Panamá no cuenta con leyes que regulen la disposición de los desechos tecnológicos, aunque sí cuenta con una ley marco, la Ley N. 33 de "Política Basura Cero" de 2018, pero falta aprobar reglamentos para su plena aplicación.

Aunque hay iniciativas, como los Anteproyectos de Ley 201 de 2018 y el 164 de 2020 (Panamá, 1018, 2020), que incentivan y promueven el reciclaje de residuos eléctricos y electrónicos para la conservación del medio ambiente, y que, por tanto, revelan la voluntad de mejora con la participación de todos los actores involucrados, la realidad es que no han dado lugar a leyes para el correcto manejo de estos desechos. Debe insistirse en que la aplicación de políticas y leyes ambientales tendientes a la recolección y reciclaje de los desechos tecnológicos es fundamental para asegurar un ciclo de vida sustentable, tanto de los aparatos eléctricos y electrónicos, como de sus residuos (Cyranek & Silva, 2010).

Para enfrentarse al problema de los desechos tecnológicos este debe enfocarse de acuerdo con cuatro ejes principales: educación, legislación, conocimiento técnico y vinculación entre los diferentes actores que intervienen en el proceso: productores y distribuidores de aparatos eléctricos y electrónicos, los propios consumidores de esos aparatos, los gestores de residuos y los gobiernos nacionales y municipales.

En esta investigación hemos centrado nuestro estudio en las empresas de desechos tecnológicos de Panamá. No debemos perder de vista que son muchos los incentivos que se relacionan con un adecuado tratamiento y reciclado de los RAEE, no solo el económico, sino también aquellos relativos a la salud pública y otros muchos relacionados con los objetivos del desarrollo sostenible, que conllevan la necesaria interacción que debe existir entre administraciones públicas, centros públicos y privados de investigación, entidades públicas, empresas privadas y otras entidades, que puedan confirmar la solidez en la transferencia y asignación de conocimiento, información y recursos en los procesos diarios de

las empresas de recuperación y reciclado de este tipo de desechos.

Nuestro objetivo ha sido profundizar en el conocimiento de las relaciones existentes entre estas empresas y las distintas entidades con las que puede interactuar en sus procesos de trabajo diario. Identificar el grado de colaboración de estas empresas con el resto de entidades permite determinar de primera mano la situación real del país, pues son parte fundamental del proceso de recolección y reciclaje de los RAEE. Las actividades de colaboración permiten a las entidades participantes obtener múltiples beneficios, entre los que se encuentran un incremento significativo de las habilidades y del conocimiento, de la calidad y fiabilidad de los productos y/o servicios que ofrecen los agentes implicados en la colaboración (López-Torres et al., 2016).

En relación con las entidades, estas los hemos clasificado en cinco tipos: empresas de la misma actividad (recolección y/o reciclaje de RAEE y/o de otros tipos de desechos), empresas de diferente actividad, instituciones públicas, organizaciones no gubernamentales sin fines de lucro y universidades públicas y/o privadas.

2. Metodología

Para obtener la información diseñamos un cuestionario con varios apartados. Aparte de las características de las empresas, nos interesaba identificar las relaciones de colaboración que estas empresas habían mantenido en los últimos cinco años con otras entidades, así como las características de dichas relaciones. Pensamos que estas relaciones pueden analizarse mediante una estructura reticular, empleando para ello dos aproximaciones diferentes: el análisis de redes sociales (ARS) y el análisis reticular de coincidencias (ARC).

2.1. Análisis de redes sociales

La aplicación de técnicas de análisis de redes sociales permite modelar la relación entre entidades como una red de nodos y arcos, esto es, un grafo, donde cada nodo representa una entidad y las relaciones entre ellas son los arcos de la red.

La aplicación del ARS se debe a que esta metodología considera la estructura de las relaciones como el patrón más valioso en la sociedad y que, a partir de su estudio directo, distingue el grado de intercambio y la intensidad en que los patrones de vinculación asignan los recursos en un sistema social (Cross et al., 2002; Gloor et al., 2008). Esas relaciones permiten el intercambio de conocimiento y de recursos entre las entidades que se relacionan, siendo este intercambio algo esencial,

de modo que una entidad está influenciada no solamente por las entidades con las que se relaciona, sino también por la red que forman las relaciones entre todas ellas (Burt, 1995).

La aplicación de técnicas de ARS para modelizar las relaciones de colaboración entre empresas se lleva realizando desde hace bastantes años. Un breve pero interesante artículo que muestra los aspectos más destacados que se pueden obtener al aplicar ARS sobre las relaciones entre empresas es (Kilkenny & Fuller-Love, 2014). Un reciente estudio bibliométrico sobre el tema (Anugerah et al., 2022) destaca las palabras clave más importantes encontradas en el estudio: innovación, gestión de proyectos, gestión del conocimiento y toma de decisiones, entre otras.

Las redes se forman, bajo la mecánica del ARS, mediante el conocimiento de las relaciones que mantienen las entidades que se van a analizar, bien preguntando a dichas entidades por esa relación, por ejemplo, utilizando un cuestionario, como es nuestro caso, bien a partir de evidencias documentales o de la observación directa. El resultado es una red, matemáticamente un grafo, donde los nodos son las entidades y los arcos las relaciones.

2.2. Análisis reticular de coincidencias

La segunda aproximación realizada en nuestro estudio proviene de la aplicación del análisis reticular de coincidencias (ARC) (Escobar, 2015). Se trata de una propuesta para el análisis estadístico de datos cuyo objetivo es captar las "pautas de concurrencia con el fin de proporcionar al investigador posibles sugerencias de cómo está estructurada la realidad" (Escobar & Tejero, 2018, p. 108). A la hora de analizar las respuestas dadas por esas empresas en el cuestionario, en relación con la colaboración con otras entidades, ha sido necesario analizar el grado de coincidencia entre ellas, para determinar aquellas que coinciden más de lo esperado por el azar. Para ello hemos utilizado el ARC.

En el ARC se parte de la idea de "escenario", que en nuestro caso es cada cuestionario rellenado por cada empresa que ha participado en el estudio, y de "suceso", que es cada posible respuesta que se da a cada pregunta del cuestionario. Bajo el enfoque del ARC los sucesos se corresponden con los nodos en una red de coincidencias, es decir, las posibles respuestas dadas a las preguntas del cuestionario que se representan en esa red. Los arcos, con su peso, serían aquellas respuestas que poseen una coincidencia dependiente, esto es, que su frecuencia de aparición conjunta sea mayor de la esperada bajo el supuesto de independencia.

Este aspecto es muy importante, pues permite diferenciar si dos sucesos no tienen nada que ver entre sí o si aparecen juntos más de lo que cabría esperar. Se puede computar un residuo entre la frecuencia esperada y la realmente observada, pudiéndose realizar un contraste de hipótesis (en este caso es de una cola, pues solo interesan sucesos que aparecen más de lo esperado por el azar) con un determinado nivel de significación dado de antemano. En nuestro caso hemos trabajado con un nivel de significación del 0,1. Aquellos nodos (sucesos) que tengan una coincidencia dependiente y que ésta sea estadísticamente significativa estarán unidos por un enlace. El peso del enlace puede darse por el valor del residuo correspondiente. Para calcular el residuo es posible utilizar varias medidas, la que nosotros hemos utilizado es el residuo de Haberman (1978), pero podrían haberse utilizado otras medidas, como la de Tanimoto, Dice, Jaccard, Rusel, Pearson o Rogers, entre otras.

Para llevar a cabo el ARC se dispone de un paquete en R, el paquete netCoin (Escobar & Martínez-Urbe, 2020), que, además de realizar todos los cálculos, permite visualizar las redes de coincidencia en un navegador web. La ventaja de esta visualización es que el investigador puede interactuar con la representación para mejorar la visualización (tamaño, color, forma y apariencia de nodos y arcos) o seleccionar uno o varios nodos de la red, esto es, las respuestas asociadas a las preguntas que interesen al investigador.

2.3. Objeto de estudio

El universo de estudio de esta investigación lo forman las empresas que se dedican a la recolección y/o reciclado de desechos tecnológicos. Encontrar las empresas fue una tarea complicada, pues se buscaba que estuvieran correctamente registradas y su principal cometido fuesen los RAEE. Para ello se recurrió al Registro de Empresas de Panamá a través de La Autoridad de la Micro, Pequeña y Mediana Empresa (AMPYME). Encontramos 13 empresas.

Ese número difiere solamente en una unidad del presentado por el PREAL (Proyectos Residuos Electrónicos en América Latina) en la conferencia "Going Green. CARE INNOVATION 2023", realizada en mayo de 2023 (Cueva et al, 2023), que encontró 14 empresas. De las 13 empresas que nosotros encontramos, solamente 9 finalizaron el estudio.

El universo de estudio fue completado con las entidades con las que dichas empresas habían mantenido relaciones de colaboración en los cinco años previos a la recogida de datos. Esa

información fue suministrada por las propias empresas en los cuestionarios que rellenaron. Esta ha sido la fuente para crear el grafo empleando el análisis de redes sociales.

El cuestionario disponía de tres apartados, perfil de la empresa, procesos de recolección y/o reciclaje y relaciones de colaboración. En el primero de ellos, además de la antigüedad, número de empleados y nivel de estudios de éstos, se preguntaba por su financiación y por las actividades de investigación, desarrollo e innovación. El segundo apartado se indagaba sobre los procesos de recolección y reciclaje en sí: forma de adquirir los desechos tecnológicos, volumen y tipos de RAEE que se manejaban, etc. El tercer apartado correspondió a las relaciones de colaboración con otras entidades, sus motivaciones, beneficios e inconvenientes encontrados en dichas colaboraciones o en la falta de ellas.

En este artículo presentamos los resultados de investigación del tercer apartado, la colaboración entre entidades, incluyendo solo información de los otros apartados cuando permitan dar explicación a los resultados obtenidos sobre colaboración. En el Anexo se incluye esta tercera parte de cuestionario.

Para la planificación del cuestionario se tomaron en cuenta datos oficiales sobre empresas de la AMPYME, bibliografía relativa a residuos electrónicos, como la de (Baldé et al., 2015; Baldé et al., 2017), las descripciones de empresas y RAEE incluidas en el R.D. 110/2015, de 20 de febrero, del Ministerio de Agricultura, Alimentación y Medio Ambiente de España, y varios estudios sobre colaboración entre entidades, como los de (Castro y Fernández, 1995; Pérez, 1996; Etkowitz y Leydesdorff, 2000; Nelson, 2008; Zazo et al., 2015; Graça & Camarinha-Matos, 2017).

Uno de los inconvenientes encontrados fue el tiempo que tomó el completado de los cuestionarios por los encargados de estas empresas, dado que resulta más difícil que los encuestados llenen la información en forma digital que cuando se efectúa persona a persona. Los últimos cuestionarios se enviaron a las empresas a principios del otoño de 2021, pero, debido a las varias oleadas de la pandemia del SARS-COVID-19, todo el proceso de recopilación y tratamiento de datos se demoró bastantes meses, hasta bien entrado 2022.

El cuestionario, en el apartado de colaboración, incluía tres grandes tipos de preguntas:

- uno, sobre la pertenencia a grupos nacionales panameños o extranjeros;

- otro, sobre las colaboraciones mantenidas con otras entidades, su número, tipos de actividades y tareas, entorno normativo, iniciativas, etc.; y,
- el tercero, sobre los resultados de la colaboración (motivaciones, valoración y grado de satisfacción, e inconvenientes encontrados). La pertenencia o no a grupos empresariales se introdujo porque en varios estudios se ha visto relación en ese aspecto y la colaboración entre empresas (Belenzon & Berkovitz, 2010; Kim & Lui, 2015; Zazo et al., 2015).

A los participantes del cuestionario no se les definió en ningún momento el concepto de colaboración. Temíamos que una definición demasiado precisa dejase fuera actividades que los propios participantes podrían entender como acciones de colaboración, con independencia de que se hubiese un acuerdo previo formalmente establecido. Más tarde, varias preguntas del cuestionario indagaban expresamente sobre qué tareas

habían desarrollado en colaboración con cada tipo de entidad.

3. Resultados y comentarios

Aplicando el ARC con el paquete netCoin se obtiene la Figura 1, en la que se muestra la red global construida a partir de las preguntas del cuestionario relativas a la colaboración con otras entidades.

La Figura 1 es una captura estática de la representación a la que puede acceder el lector en <http://fenix.usal.es/relaciones/> para visualizar la red y, sobre todo, interaccionar con ella. Cada pregunta y sus respuestas asociadas se representan con un pequeño círculo del mismo color. El primero de ellos, de color verde, representa dos preguntas agrupadas en una sola, a saber (i) si la empresa había mantenido relaciones de colaboración con otras entidades [sí / no]; y (ii) cómo se había canalizado esa colaboración [improvisación / tarea adicional / existe departamento encargado / depende de la colaboración].



Figura 1. Captura de pantalla de la Red global obtenida con netCoin para las relaciones de colaboración de las empresas encuestadas. Los números que aparecen delante de las respuestas se corresponden con el número de pregunta en el cuestionario, que se puede ver en el Anexo. Véase <http://fenix.usal.es/relaciones/>

El círculo de color rojo de la Figura 1 se corresponde con la pregunta de si se había colaborado o no con entidades extranjeras (sin considerar el tipo de entidad). El tercer círculo, de color rosa, consistía en una pregunta de respuesta múltiple, pues se permitía en el cuestionario marcar varios tipos de entidades panameñas con las que se había colaborado en los últimos cinco años: empresas misma actividad / empresas distinta actividad

/ entidades públicas / ONG / Universidades públicas / Universidades privadas.

El cuarto y quinto círculo, de colores naranja y azul, respectivamente, están relacionados con preguntas sobre pertenencia o no a grupo panameño o extranjero de empresas. Estas preguntas tienen interés para identificar el grado de coincidencia de las respuestas correspondientes y las respuestas dadas al resto de preguntas.

La importancia de la presentación indicada en la Figura 1 (se recomienda visitar el enlace previo) radica en que se pueden seleccionar y filtrar nodos en la leyenda de la parte superior derecha. Así, si se filtra el nodo "Colaboración", se obtiene una red en la que aparecen los nodos correspondientes a las preguntas sobre colaboración (sí/no) y tipo de entidad, junto con nodos de otras preguntas con las que comparte enlace. La visualización de la red de la Figura 1 nos informa de varios aspectos interesantes.

En primer lugar, resaltar el hecho de que si existe un enlace entre nodos es que existe coincidencia dependiente entre respuestas del cuestionario y ésta es estadísticamente significativa al nivel de significación del 0,1. El grosor de cada enlace muestra el valor del residuo asociado entre respuestas, siendo, cuanto más grueso, más significativo.

Podemos ver en ese grafo que la respuesta "3.13. Canalización: Improvisación" aparece como nodo aislado. Esta respuesta se corresponde con la pregunta sobre cómo se gestiona la colaboración con otras entidades. Que aparezca en el grafo nos indica que esa respuesta se ha dado por alguna de las empresas que respondieron el cuestionario, pero al estar aislada significa que esa respuesta no tiene coincidencia dependiente y significativa con las respuestas dadas al resto de preguntas. Si se hubiese ampliado el nivel de significación (un intervalo de confianza menor) seguramente esa respuesta tendría enlace con otras, aunque su residuo sería muy pequeño. Es decir, tendría dependencia dependiente, pero no ser significativa al nivel elegido. Lo mismo ocurre con la respuesta "3.31.Univ. Públicas: No" que también se encuentra aislado en el grafo.

Otro aspecto que deseamos destacar es que se ha incluido en el estudio las respuestas negativas "no se ha colaborado con" y "no pertenece a" (nodos "3.34. Univ.Privadas: No", "3.19. EmpresaDistintaAct: No" y otros muchos). Ha sido intencionado, pues nos interesa detectar respuestas con posibles coincidencias dependientes entre respuestas negativas y otras afirmativas. Sin embargo, tal como vemos en el grafo las respuestas afirmativas solo aparecen enlazadas entre sí, al igual que las respuestas negativas también aparecen solo enlazadas entre sí, salvo la rama del componente de la izquierda del grafo, que enlaza un par de nodos afirmativos. Excepto esa rama, podemos apreciar que existe una simetría casi completa entre componentes afirmativos y negativos del grafo.

Debido a este hecho, los resultados los realizaremos en el sentido de las respuestas afirmativas,

que son las que muestran las relaciones de colaboración. Así, el componente de respuestas afirmativas formado por dos nodos "3.7. Colaboración: Sí" y "3.15. EmpresasMismaAct.: Sí" que aparece en la parte superior del grafo, simétrico respecto de las respuestas negativas por encima de él, nos lleva a afirmar que las empresas encuestadas han mantenido relaciones de colaboración especialmente con otras empresas que también se dedican a la recolección y/o reciclaje de desechos.

Decíamos que el componente de la izquierda aparecía una rama con enlaces entre los nodos "3.31. Univ.Pública: Sí", "3.13. Canalización: Depto" y "3.27. ONG:No", ello se debe a que se ha dado la coincidencia que una de las empresas no se relaciona con ONG, aunque sí con otras, y es la única empresa que dispone de un departamento concreto para las labores de colaboración con otras entidades.

En cuanto al componente de respuestas afirmativas del grafo que aparece a la derecha, la clave aquí está en entender que las relaciones deben verse como enlaces dos a dos, esto es coincidencias en las respuestas asociadas del cuestionario. El enlace existente entre las respuestas "3.34. Univ.Privadas: Sí" y "3.19. EmpresaDistintaAct: Sí" se debe a que dos empresas, y solo ellas, han colaborado con universidades privadas, además de que también lo han hecho con empresas que no se dedican a la recolección y/o reciclado de desechos. Por eso la coincidencia es significativa. Esas dos empresas han colaborado con ambas entidades y, efectivamente, al igual que para el ARS se presupone un flujo de información o de saber hacer entre nodos (Burt, 1995), de ahí que pensemos que ese flujo de información también puede tenerse entre entidades con el ARC, aunque siempre teniendo en cuenta cuáles las entidades que responden los cuestionarios.

En ese componente vemos también el enlace entre "3.19.EmpresaDistintaAct: Sí" y "3.23. Inst.Públicas: Sí". Ello se debe a que otras dos empresas han mantenido colaboración con instituciones públicas, solo esas dos, pero también con empresas de actividad diferente a recolección y reciclado de desecho. Destacamos que no tienen por qué ser las mismas empresas que el caso anterior, de hecho, coincide solo una de las empresas con el enlace explicado en el párrafo anterior.

Queremos también explicar los enlaces entre "3.34. Univ.Privadas: Sí" y "3.13. Canalización: Tarea Adicional", y este enlace con "3.23. Inst.Públicas: Sí". En este caso estos enlaces se deben a que una empresa de las encuestadas

El resto de las empresas de desechos tecnológicos que respondieron el cuestionario manifestó que, de media, solo mantenía relaciones con aproximadamente dos entidades, alguna de ellas solo con universidades. Uno de los motivos puede ser la juventud de algunas de ellas, precisamente las que menos relaciones tienen. Otro motivo es que varias de las empresas poseen muy pocos empleados: tres de ellas poseen menos de 5 empleados y otra menos de 10. Una de las empresas manifestó no mantener relaciones, por eso aparece como un nodo aislado (DT9).

En el grafo de la Figura 2 se aprecia que el color rojo es el predominante, lo cual se corresponde con que las relaciones tienden a suceder entre entidades que se dedican a la recolección y/o de varios tipos de desechos, no solo los tecnológicos. Las actividades que suelen realizarse por medio de esa colaboración se centran en intercambio y comercialización de materiales reciclables, confirmación de precios de mercado, logística de residuos y coordinación de gestión de residuos para un mismo cliente.

Los nodos que aparecen en color magenta son empresas dedicadas a actividades diferentes a las de recolección y/o reciclado de residuos. Se aprecia que solo tres de las empresas encuestadas habían mantenido colaboración con ese tipo de entidades, destacando, de nuevo DT2, que mantiene relación con hasta cinco empresas de otras actividades. Las otras dos son DT5 y DT8, que colaboran solamente con una empresa de otras actividades.

En relación con las empresas dedicadas a otras actividades, la colaboración con ellas suele plasmarse en la gestión de residuos, incluidos la separación de componentes, servicios de destrucción y compra de materiales, también el tratamiento de desechos especiales y, algunas veces, concienciar sobre los desechos tecnológicos. Las tareas que suelen realizarse en colaboración con las universidades también pasan por la gestión de residuos y la concienciación sobre responsabilidad social de éstos. Las actividades con las instituciones públicas se plasman en participación en seminarios y proyectos especiales y en la gestión de residuos. Las actividades que se realizan con ONG se plasman en participación en eventos de reciclaje, apoyo y capacitación en proyectos de gestión de residuos y ayuda en separación de componentes.

Llama la atención que ninguna entidad referenciada lo haya sido por más de una empresa. Sin duda eso se debe al bajo número de relaciones que mantienen, en general, las empresas de desechos tecnológicos que respondieron al cuestionario. Podría argumentarse que dichas

empresas tienen ámbitos regionales diferentes, o que se encuentran en regiones de Panamá apartadas entre sí, sin embargo, no es así, casi todas ellas están en la provincia de Panamá y solo una en la provincia de Panamá Oeste.

En cuanto a la pertenencia a grupos nacionales panameños o extranjeros de las empresas, los cuestionarios muestran que ese aspecto no influye en el número de entidades con las que las empresas se relacionan, al contrario de lo que en algunos casos se ha encontrado (Belenzon & Berkovitz, 2010; Kim & Lui, 2015). Así, DT2 no pertenece a ningún grupo nacional o extranjero, DT8 pertenece a ambos tipos y DT5 solo a grupos nacionales.

Preguntadas las empresas sobre los principales inconvenientes en relación con la falta de colaboración con otras entidades, estas manifestaron que el mayor inconveniente era la poca divulgación sobre reciclado de desechos tecnológicos que tienen, no solo la comunidad, sino también, de manera destacada, otras entidades, algunas de ellas con importante responsabilidad en el tema, como son las instituciones públicas. Añadieron además la falta de leyes y normativas, la excesiva burocracia, la falta de personal que se dedique de manera expresa a establecer relaciones de colaboración con otras entidades o la falta de incentivos económicos para realizar colaboraciones. Algunos otros inconvenientes mencionados fueron que, al presentar sus servicios, no son respondidas las cartas, correos, ni llamadas en las entidades gubernamentales pese a haber anunciados próximos descartes. También que existe poco interés en crear conglomerados para la atención de clientes de forma integral.

En relación con las motivaciones que llevaron a las empresas encuestadas a colaborar con otras entidades, se destaca la de obtener información, experiencia o materiales necesarios para el desarrollo de los objetivos de la empresa. También fue importante la motivación de ampliar relaciones profesionales y mejorar la imagen y prestigio de la empresa. En general, el grado de satisfacción de las empresas encuestadas con el resultado de la relación, en relación con estas motivaciones, fue alto, lo cual supone una clara aspiración de esas empresas por colaborar y adquirir conocimientos.

4. Conclusiones

En este artículo hemos descrito los resultados de una investigación exploratoria sobre cómo se relacionan las empresas que se dedican a la recolección y/o reciclado de residuos de aparatos eléctricos y electrónicos en la República de Pa-

namá. Son pocas las empresas operativas existentes y menos las que han decidido participar en la investigación. Pocos estudios existen a este respecto en el contexto de América Latina y El Caribe.

A través del cuestionario y de las respuestas dadas a este hemos aplicado dos técnicas de análisis reticular, como es el ARS y el ARC para conocer cómo se relacionan estas empresas con otras entidades. En el análisis que hemos hecho empleando ARS no hemos incluido medidas típicas de grafo, como son las medidas de centralidad, ni tampoco hemos hecho otros análisis que sí se realizan en esos estudios. El motivo de no haberlo hecho no es otro que el número tan bajo de relaciones que existen en el grafo. El resultado más destacado es que solamente dos de las empresas encuestadas mantienen un alto grado de relaciones con otras entidades, para el resto el número de relaciones es pequeño. Las pocas relaciones obtenidas aplicando el ARS nos muestran que la transferencia de conocimiento, información y recursos entre las entidades, que sería deseable, es, en realidad, muy escasa. Es este un aspecto negativo que debiera mejorarse para potenciar el sistema de innovación asociado al sector.

Tampoco hemos encontrado relación entre la pertenencia a grupos nacionales o extranjeros y el número de relaciones que las empresas mantienen con otras entidades.

Sorprende el hecho de que en varias ocasiones algunas empresas participantes en el estudio hayan indicado que han colaborado con empresas que también participaron en el mismo, pero estas últimas no referenciaron a las primeras. Ello sugiere, principalmente, dos posibles escenarios: uno, que dichas empresas han omitido expresamente esa información por tener algún tipo de recelo en darla; el segundo sería aquel en el que, aunque hayan colaborado, puede que lo hayan hecho sin existir una relación formal y previamente establecida, de modo que unas empresas sí la hayan indicado en el cuestionario y otras no.

Las tareas que más se han dado en las relaciones con empresas dedicadas a la recolección y/o reciclaje (de todo tipo de residuos) son las de intercambio y comercialización de materiales y logística de residuos. Con otras empresas las tareas han sido las propias de gestión de residuos, separación de componentes, destrucción de materiales y tratamiento de desechos especiales. Con entidades públicas y universidades las tareas se han centrado en la gestión de residuos y la concienciación sobre responsabilidad social del tratamiento de estos, siendo esta última actividad la mayormente realizada con ONG. Destacamos la implicación de las propias empresas de RAEE en

labores de concienciación social sobre recolección y reciclado de estos tipos de residuos.

La aplicación del ARC nos ha permitido analizar las respuestas dadas en el cuestionario desde el punto de vista de las coincidencias estadísticamente significativas. Sin embargo, en nuestro caso, el número de empresas que rellenaron cuestionario fue pequeño, eso hace que se hayan obtenido coincidencias significativas con solamente una o dos respuestas coincidentes. Este ha sido el mayor inconveniente que hemos encontrado al aplicar el ARC. Pese a ello, los resultados son consistentes y estadísticamente robustos.

Las relaciones de colaboración contribuyen a generar intercambio de información, conocimientos y de recursos entre las entidades que colaboran. Para las empresas panameñas dedicadas a la recolección y al reciclaje de desechos tecnológicos es preponderante que estas relaciones de colaboración existan y se fomenten para aprovechar la oportunidad que el tratamiento de este tipo de desechos ofrece en la generación de riqueza y de muchos puestos de trabajo y, con especial atención, a concientizar a la comunidad en general acerca de sus efectos y a la determinación de medidas concretas que puedan minimizar la contaminación ambiental que estos desechos producen a nivel local, regional y mundial. Sin embargo, la escasa divulgación, la falta de leyes y normas sobre desechos tecnológicos, la poca organización y excesiva burocracia en los procesos operativos y administrativos, no favorece esas relaciones. A este respecto, es interesante la investigación de (Farnum & Kelly, 2019) en la que se caracterizan, dentro del proceso de recuperación y reciclaje, a los recicladores de oficio en Panamá, y en la que, lamentablemente, se expone la poca vinculación entre estos trabajadores y las instituciones públicas, y la necesidad de planes de gestión integral de residuos que puedan servir de plataforma para la creación de políticas públicas inclusivas, basadas en la ciencia y la innovación social. Es necesario que aumente la divulgación sobre los beneficios que la recolección y reciclado de los RAEE tiene sobre la población y la salud ambiental.

La escasa concienciación que existe y se divulga de manera muy limitada por las empresas encuestadas, por ONG relacionadas con el tema y también por algunas instituciones públicas, debiera ampliarse y extenderse a todo el país desde las más altas esferas, mediante la creación de leyes que regulen de manera concreta este tipo de desechos. Es necesario que se regulen los procesos técnicos y las vinculaciones necesarias entre los diferentes actores que intervienen en el proceso de recolección y reciclado de los RAEE,

como son los productores y distribuidores de aparatos eléctricos y electrónicos, los propios consumidores de esos aparatos, las instituciones públicas, los gestores de residuos, el gobierno nacional y los responsables regionales y municipales.

Los procedimientos administrativos que se derivan de esas leyes deben simplificarse todo lo posible, eliminando la excesiva burocracia que pudiera existir, siendo este aspecto uno de los inconvenientes manifestado por algunas de las empresas encuestadas. Debe potenciarse la colaboración entre empresas de desechos tecnológicos, que permitan agilizar los procesos administrativos, y que se dote de personal especializado que facilite la colaboración entre los diferentes actores involucrados, así como incentivos económicos para desarrollar dichas colaboraciones y que se potencie la inversión en los diferentes procesos que intervienen en la recolección y reciclado de los desechos tecnológicos.

Agradecimientos

Este trabajo se ha realizado dentro del proyecto “Sistema de Redes de las empresas de desechos tecnológico en Panamá”, código VIP-01-17-01-2020-01, de la Universidad de Panamá.

Anexo

Cuestionario II: Relaciones de colaboración de la empresa

3. Información de las relaciones

En este apartado se le preguntará sobre relaciones que colaboración (en un sentido muy amplio) que haya podido mantener su empresa con las siguientes *entidades panameñas*:

- Empresas de su misma actividad.
- Empresas de otras actividades.
- Instituciones públicas (municipios, secretarías, ministerios, escuelas, institutos de enseñanza media, etc.).
- Organizaciones no gubernamentales sin fines de lucro (ONGs, fundaciones, etc.).
- Universidades públicas o privadas (incluidos institutos universitarios, centros y laboratorios científicos de estas universidades).

3.1 ¿Su empresa forma parte de algún *grupo panameño* de empresas (de su misma actividad o de actividades diferentes)?

- Sí (Pase a la pregunta 3.2)
- No (Pase a la pregunta 3.4)

3.2 Esta relación con algún *grupo panameño* de empresas surge a partir de: (Puede marcar varias respuestas)

- Imperativo legal o normativo
- Interés de la empresa
- Otro. Especificar:

3.3 Especifique a cuántos *grupos panameños* de empresas pertenece su empresa:

Número de grupos panameños:

3.4 ¿Su empresa forma parte de algún *grupo extranjero* de empresas (de su misma actividad o de actividades diferentes)?

- Sí (Pase a la pregunta 3.5)
- No (Pase a la pregunta 3.7)

3.5 Esta relación con algún *grupo extranjero* de empresas surge a partir de: (Puede marcar varias respuestas)

- Imperativo legal o normativo
- Interés de la empresa
- Otro. Especificar:

3.6 Especifique a cuántos *grupos extranjeros* de empresas pertenece su empresa:

Número de grupos extranjeros:

3.7 Durante los últimos 5 años, ¿la empresa ha mantenido relaciones de colaboración con alguna de las siguientes *entidades panameñas*: otras empresas de su misma actividad, empresas de otras actividades, instituciones públicas, organizaciones no gubernamentales sin fines de lucro, o con universidades públicas o privadas?:

- Sí (Pase a la pregunta 3.13)
- No (Pase a la pregunta 3.8)

3.8 A pesar de no haber mantenido relaciones de colaboración con *entidades panameñas*, ¿podría indicarnos si pudieran haber sido de interés para su empresa?

- De mucho interés
- De bastante interés
- De poco interés
- De ningún interés

3.9 Podría indicarnos los inconvenientes que ha encontrado su empresa para no mantener relaciones de colaboración con *entidades panameñas*: (Puede marcar varias opciones)

- La empresa no dispone de personal que se encargue de dicha tarea
- La empresa no cuenta con infraestructura tecnológica
- La empresa no dispone de apoyo gubernamental para hacerlo
- Lo hemos querido hacer pero no hemos encontrado el momento de hacerlo
- Lo hemos hecho en el pasado, pero en los últimos 5 años no lo hemos concretado
- Hay poca información para encontrar entidades adecuadas a los objetivos de la empresa
- Hay diferencias entre las prácticas de trabajo de la empresa y las otras entidades
- Hay insuficiente regulación en materia de desechos tecnológicos para poder establecer relación con otras entidades
- Hay poca divulgación sobre desechos tecnológicos para poder establecer relación con otras entidades
- Hay poca divulgación a la comunidad sobre a los desechos tecnológicos para que concienticen su importancia
- Las negociaciones y procesos administrativos son muy complicados
- El débil fomento de leyes que permitan establecer relaciones de colaboración con otras entidades
- La financiación pública disponible
- El apoyo económico de las empresas de otras actividades
- Otro. Especificar:

3.10 En caso de que a la empresa le *hubiera gustado* mantener relaciones de colaboración con *entidades panameñas*, por favor, indique por orden de importancia *las entidades con las que le hubiera gustado colaborar*: (Califique de 1 a 6, donde 1 es de mayor importancia y 6 es de menor importancia)

- Empresas de su misma actividad
- Empresas de otras actividades
- Instituciones públicas
- Organizaciones no gubernamentales sin fines de lucro
- Universidades públicas y particulares
- Otro. Especificar:

3.11 En caso de que a la empresa le *hubiera gustado* mantener relaciones de colaboración con *entidades panameñas*, por favor, indique *qué motivaciones podría tener su empresa*: (Puede marcar varias opciones)

- Conseguir ayuda financiera para que su empresa crezca
- Formar parte de una red profesional o ampliar relaciones profesionales
- Obtener información, experiencia o material necesarios para el desarrollo de los objetivos de la empresa
- Mejorar la imagen y prestigio de la empresa
- Acceder a instalaciones o equipamiento no disponible en su empresa
- Acceder a nuevas relaciones de colaboración
- Asesorar o participar para crear o hacer viable una nueva empresa
- Beneficio económico
- Acceder a una patente
- Otro. Especificar:

3.12 Conteste esta pregunta solamente si su empresa ha mantenido relaciones de colaboración con *entidades extranjeras*. En términos generales, indique las motivaciones de su empresa que han permitido mantener relaciones de colaboración con *entidades extranjeras*, tales como empresas de su misma actividad, empresas de otras actividades, instituciones públicas, organizaciones no gubernamentales sin fines de lucro, o con universidades públicas o privadas: (Puede marcar varias respuestas)

- Conseguir ayuda financiera para que la empresa crezca
- Formar parte de una red profesional o ampliar relaciones profesionales

Obtener información, experiencia o material necesarios para el desarrollo de los objetivos de la empresa

Mejorar la imagen y prestigio de la empresa

Acceder a instalaciones o equipamiento no disponible en la empresa

Acceder a nuevas relaciones de colaboración

Asesorar o participar para crear o hacer viable una nueva empresa

Beneficio económico

Acceder a una patente

Otro. Especificar:

Ha finalizado el cuestionario, ¡gracias por su apoyo! Si tiene cualquier otra información que añadir relacionada con este estudio, o bien algún comentario referido a su relación con otras entidades, por favor, hágalo a continuación. ¡Muchas gracias por su colaboración!

3.13 La gestión de las relaciones de colaboración de la empresa con *entidades panameñas* (con las empresas de su misma actividad, empresas de otras actividades, instituciones públicas, organizaciones no gubernamentales sin fines de lucro, o con universidades públicas o privadas) se canaliza a través de:

Un departamento o sección específica de la empresa

Un personal que además de cumplir otras funciones, debe hacerse cargo de la gestión de las relaciones

La improvisación de acuerdo con la necesidad de colaboración

La contratación de otros para que se encarguen de la gestión de las relaciones

Otro. Especificar:

3.14 Indique por orden de importancia las *entidades panameñas* con las que la empresa desearía seguir manteniendo o establecer nuevas relaciones de colaboración: (Califique de 1 a 6, donde 1 es de mayor importancia y 6 es de menor importancia)

Empresas de su misma actividad

Empresas de otras actividades

Instituciones públicas

Organizaciones no gubernamentales sin fines de lucro

Universidades públicas y particulares

Otro. Especificar:

3.15 ¿La empresa ha mantenido relaciones de colaboración con *empresas panameñas de su misma actividad* durante los 5 últimos años?

Sí (Pase a la pregunta 3.16)

No (Pase a la pregunta 3.19)

3.16 Indique el número de las *empresas panameñas de su misma actividad* con las que ha mantenido relaciones de colaboración durante los 5 últimos años:

Menor a 5 empresas de su misma actividad

Entre 5 y 9 empresas de su misma actividad

Entre 10 y 15 empresas de su misma actividad

Entre 15 y 20 empresas de su misma actividad

Más de 20 empresas de su misma actividad

3.17 Por favor ordene el nombre de las *empresas panameñas de su misma actividad* que con mayor frecuencia ha mantenido relaciones de colaboración durante los 5 últimos años:

3.18 Podría indicar de manera breve en qué tareas ha colaborado con *otras empresas panameñas de su misma actividad*:

3.19 ¿La empresa ha mantenido relaciones de colaboración con las *empresas panameñas de otras actividades* durante los 5 últimos años?

Sí (Pase a la pregunta 3.20)

No (Pase a la pregunta 3.23)

3.20 Indique el número *empresas panameñas de otras actividades* que ha mantenido relaciones de colaboración durante los 5 últimos años:

Menor a 5 empresas de otras actividades

Entre 5 y 9 empresas de otras actividades

Entre 10 y 15 empresas de otras actividades

Entre 15 y 20 empresas de otras actividades

Más de 20 empresas de otras actividades

3.21 Por favor ordene el nombre de las *empresas panameñas de otras actividades* que con mayor frecuencia ha mantenido relaciones de colaboración durante los 5 últimos años:

3.22 Podría indicar de manera breve en qué tareas ha colaborado con esas *empresas panameñas de otras actividades*:

3.23 ¿La empresa ha mantenido relaciones de colaboración con las *instituciones públicas panameñas* durante los 5 últimos años?

Sí (Pase a la pregunta 3.24)

No (Pase a la pregunta 3.27)

3.24 Indique el número de las *instituciones públicas panameñas* que ha mantenido relaciones de colaboración durante los 5 últimos años:

Menor a 5 instituciones públicas

Entre 5 y 9 instituciones públicas

Entre 10 y 15 instituciones públicas

Entre 15 y 20 instituciones públicas

Más de 20 instituciones públicas

3.25 Por favor ordene el nombre de las *instituciones públicas panameñas* que con mayor frecuencia ha mantenido relaciones de colaboración durante los 5 últimos años:

3.26 Podría indicar de manera breve en qué tareas ha colaborado con esas *instituciones públicas panameñas*:

3.27 ¿La empresa ha mantenido relaciones de colaboración con las *organizaciones no gubernamentales sin fines de lucro panameñas* durante los 5 últimos años?

Sí (Pase a la pregunta 3.28)

No (Pase a la pregunta 3.31)

3.28 Indique el número de las *organizaciones no gubernamentales sin fines de lucro panameñas* con las que ha mantenido relaciones de colaboración durante los 5 últimos años:

Menor a 5 organizaciones no gubernamentales sin fines de lucro

Entre 5 y 9 organizaciones no gubernamentales sin fines de lucro

Entre 10 y 15 organizaciones no gubernamentales sin fines de lucro

Entre 15 y 20 organizaciones no gubernamentales sin fines de lucro

Más de 20 organizaciones no gubernamentales sin fines de lucro

3.29 Por favor ordene el nombre de las *organizaciones no gubernamentales sin fines de lucro panameñas* que con mayor frecuencia ha mantenido relaciones de colaboración durante los 5 últimos años:

3.30 Podría indicar de manera breve en qué tareas ha colaborado con esas *organizaciones no gubernamentales sin fines de lucro panameñas*:

3.31 ¿La empresa ha mantenido relaciones de colaboración con las *universidades públicas panameñas* durante los 5 últimos años?

Sí (Pase a la pregunta 3.32)

No (Pase a la pregunta 3.34)

3.32 Indique las *universidades públicas panameñas* con las que ha mantenido relaciones de colaboración durante los 5 últimos años: (Por favor, ordene por frecuencia: 1 con la que más ha colaborado, luego 2, etc.)

Universidad de Panamá

Universidad Tecnológica de Panamá

Universidad Autónoma de Chiriquí

Universidad Especializada de las Américas

Universidad Marítima Internacional de Panamá

3.33 Podría indicar de manera breve en qué tareas ha colaborado con esas *universidades públicas panameñas*:

3.34 ¿La empresa ha mantenido relaciones de colaboración con las *universidades privadas panameñas*?:

Sí (Pase a la pregunta 3.35)

No (Pase a la pregunta 3.38)

3.35 Indique el número de *universidades privadas panameñas* con las que ha mantenido relaciones de colaboración durante los 5 últimos años:

Menor a 5 universidades privadas

Entre 5 y 9 universidades privadas

Entre 10 y 15 universidades privadas

Entre 15 y 20 universidades privadas

Más de 20 universidades privadas

3.36 Por favor ordene el nombre de las *universidades privadas panameñas* que con mayor frecuencia ha mantenido relaciones de colaboración durante los 5 últimos años:

3.37 Podría indicar de manera breve en qué tareas ha colaborado con esas *universidades privadas panameñas*:

Ha finalizado esta sección del cuestionario, ¡gracias por su apoyo!

4. Motivaciones, grado de satisfacción y obstáculos en la colaboración

4.1 Marque las motivaciones como Sí/No, durante los últimos 5 años, que han podido impulsar a su empresa a establecer y mantener relaciones de colaboración con otras *entidades panameñas*: empresas de su misma actividad, empresas de otras actividades, instituciones públicas, organizaciones no gubernamentales sin fines de lucro, o con universidades públicas o privadas. A continuación, valore el grado de satisfacción que han tenido para su empresa. (Valore la satisfacción únicamente para las motivaciones que haya marcado previamente "Sí"):

#	Motivaciones	Sí	No	Grado de satisfacción para su empresa			
				Ninguno	Poco	Bastante	Mucho
1	Conseguir ayuda financiera para que la empresa crezca	x	x	1	2	3	4
2	Formar parte de una red profesional o ampliar relaciones profesionales	x	x	1	2	3	4
3	Obtener información, experiencia o material necesarios para el desarrollo de los objetivos de la empresa	x	x	1	2	3	4
4	Mejorar la imagen y prestigio de su empresa	x	x	1	2	3	4
5	Acceder a instalaciones o equipamiento no disponible en su empresa	x	x	1	2	3	4
6	Acceder a nuevas relaciones de colaboración	x	x	1	2	3	4
7	Asesorar o participar para crear o hacer viable una nueva empresa	x	x	1	2	3	4
8	Acceder a una patente	x	x	1	2	3	4
9	Beneficio económico	x	x	1	2	3	4
10	Otros (especificar):			1	2	3	4

4.2 Señale los posibles inconvenientes que pueden impedir las relaciones de colaboración entre su empresa y otras *entidades panameñas*: (Puede marcar varias respuestas)

- La empresa no dispone de personal que se encargue de dicha tarea
- La empresa no cuenta con infraestructura tecnológica
- La empresa no dispone de apoyo gubernamental para hacerlo
- Lo hemos querido hacer pero no hemos encontrado el momento de hacerlo
- Lo hemos hecho en el pasado, pero en los últimos 5 años no lo hemos concretado
- Hay poca información para encontrar entidades adecuadas a los objetivos de la empresa
- Hay diferencias entre las prácticas de trabajo de la empresa y las otras entidades
- Hay insuficiente regulación en materia de desechos tecnológicos para poder establecer relación con otras entidades
- Hay poca divulgación en lo que se refiere a los desechos tecnológicos para poder establecer relación con otras entidades
- Hay poca divulgación a la comunidad sobre los desechos tecnológicos para que concienticen su importancia
- Las negociaciones y procesos administrativos son muy complicados
- El débil fomento de leyes que permitan establecer relaciones de colaboración con otras entidades
- La financiación pública disponible
- El apoyo económico de las empresas de otras actividades
- Otro. Especificar:

4.3 Si su empresa ha mantenido relaciones de colaboración con *entidades extranjeras* (las empresas de su misma actividad, empresas de otras actividades, instituciones públicas, organizaciones no gubernamentales sin fines de lucro y universidades públicas o privadas), puede valorar que la colaboración ha sido mejor que con *entidades panameñas*:

- Sí (Pase a la pregunta 4.4)
- No (Ha finalizado el cuestionario, ¡gracias por su apoyo! Si lo desea puede rellenar el cuadro final)

4.4 En términos generales, indique las motivaciones de su empresa que han permitido mantener relaciones de colaboración con *entidades extranjeras*: (Puede marcar varias respuestas)

- Conseguir ayuda financiera para que la empresa crezca
- Formar parte de una red profesional o ampliar relaciones profesionales
- Obtener información, experiencia o material necesarios para el desarrollo de los objetivos de la empresa
- Mejorar la imagen y prestigio de la empresa
- Acceder a instalaciones o equipamiento no disponible en la empresa
- Acceder a nuevas relaciones de colaboración
- Asesorar o participar para crear o hacer viable una nueva empresa
- Beneficio económico
- Acceder a una patente
- Otro. Especificar:

Ha finalizado el cuestionario, ¡gracias por su apoyo! Si tiene cualquier otra información que añadir relacionada con este estudio, o bien algún comentario referido a su relación con otras entidades, por favor, hágalo a continuación. ¡Muchas gracias por su colaboración!

Referencias

AAUD; INECO. (2017). Plan Nacional de Gestión Integral de Residuos 2017-2027. Autoridad del Aseo Urbano y Doméstico e Ingeniería y Economía del Transporte.

- Anugerah, A. R.; Muttaqin, P. S.; Trinarningsih, W. (2022). Social network analysis in business and management research: A bibliometric analysis of the research trend and performance from 2001 to 2020. *Heliyon*, 8:4, e09270. <https://doi.org/10.1016/j.heliyon.2022.e09270>
- Baldé, C. P.; Forti, V.; Gray, V.; Kuehr, R.; Stegmann, P. (2017). The Global E-waste Monitor—2017: Quantities, Flows and Resources. United Nations University, International Telecommunication Union, and International Solid Waste Association.
- Baldé, C. P.; Blumenthal, K.; Fondeur Gill, S.; Huisman, J.; Kern, M.; Kuehr, R.; Magpantay, E.; Micheli, E. P. (2015). E-waste statistics: Guidelines on classification, reporting and indicators 2015. United Nations University, IAS - SCYCLE, Bonn, Germany.
- Belenzon, S.; Berkovitz, T. (2010). Innovation in Business Groups. // *Management Science*. 56:3, pp. 519-35. JSTOR. <https://www.jstor.org/stable/27784131>
- Burt, R. S. (1995). *Structural Holes: The Social Structure of Competition*. Harvard University Press, Cambridge, Ma.
- Castro E.; y Fernández, I. (1995). La nueva política de articulación del Sistema de Innovación en España. Concepción, Chile.
- Cross, R.; Borgatti, S. P.; Parker, A. (2002). Making Invisible Work Visible: Using Social Network Analysis to Support Strategic Collaboration. // *California Management Review*. 44:2, 25-46. <https://doi.org/10.2307/41166121>
- Cueva, A.; Hernandez, C.; Alhilali, S.; Ives-Keeler, K.; Casasnas, B. (2023) Promoting circular economy through resource-efficient electronic recycling across Latin America. // International CARE Electronics Office. "Going Green - CARE INNOVATION 2023 Conference Program & Abstract Book". 9-11 mayo. Vienna, Austria.
- Cyranek, G.; Silva, U. (2010). Los residuos electrónicos: Un desafío para la Sociedad del Conocimiento en América Latina y el Caribe. Montevideo: UNESCO, Plataforma RELAC SUR/IDRC.
- Escobar, M. (2015). Studying Coincidences with Network Analysis and Other Multivariate Tools. // *The Stata Journal*. 15:4, 1118-1156. <https://doi.org/10.1177/1536867X1501500410>
- Escobar, M.; Martínez-Urbe, L. (2020). Network Coincidence Analysis: The netCoin R Package. *Journal of Statistical Software*. 93, 1-32. <https://doi.org/10.18637/jss.v093.i11>
- Escobar, M.; Tejero, C. (2018). El análisis reticular de coincidencias. // *Empiria: Revista de metodología de ciencias sociales*. 39, Article 39. <https://doi.org/10.5944/empiria.39.2018.20879>
- España. Real Decreto 110/2015, de 20 de febrero, sobre residuos de aparatos eléctricos y electrónicos. Ministerio de Agricultura, Alimentación y Medio Ambiente. // *BOE*. 45 (21 de febrero). <https://www.boe.es/eli/es/rd/2015/02/20/110/con>.
- Etzkowitz, H.; y Leydesdorff, L. (2000). The dynamics of innovation: from National Systems and "mode 2" to a Triple Helix of university-industry-government relations. // *Research Policy*. 29:2 (2000), 109-123. [https://doi.org/10.1016/S0048-7333\(99\)00055-4](https://doi.org/10.1016/S0048-7333(99)00055-4)
- Farnum, F.; Kelly, R. (2019). Primera caracterización nacional de recicladores de oficio de Panamá. // *Revista Saberes APUDEP*. 2:1, 59-79.
- Forti, V.; Balde, C. P.; Kuehr, R.; Bel, G. (2020). The Global E-waste Monitor 2020: Quantities, flows and the circular economy potential. United Nations University/United Nations Institute for Training and Research, International Telecommunication Union, and International Solid Waste Association. <https://collections.unu.edu/view/UNU:7737>

- Gloor, P. A.; Paasivaara, M.; Schoder, D.; Willems, P. (2008). Finding collaborative innovation networks through correlating performance with social network structure. // *International Journal of Production Research*. 46:5, 1357-1371. <https://doi.org/10.1080/00207540701224582>
- Graça, P.; Camarinha-Matos, L. M. (2017). Performance indicators for collaborative business ecosystems—Literature review and trends. // *Technological Forecasting and Social Change*. 116, 237-255. <https://doi.org/10.1016/j.techfore.2016.10.012>
- Haberman, S. J. (1978). *Analysis of Qualitative Data* (Vol. 1). Academic Press.
- Kilkenny, M.; Fuller-Love, N. (2014). Network analysis and business networks. // *International Journal of Entrepreneurship and Small Business*. 21:3, 303-316. <https://doi.org/10.1504/IJESB.2014.060894>
- Kim, Y. Lui, S.S. (2015). The impacts of external network and business group on innovation: Do the types of innovation matter? // *Journal of Business Research*. 68:9, 1964-1973. <https://doi.org/10.1016/j.jbusres.2015.01.006>
- López-Torres, G. C.; Maldonado Guzmán, G.; Pinzón Castro, S. Y.; García Ramírez, R.; López-Torres, G. C.; Maldonado Guzmán, G.; Pinzón Castro, S. Y.; García Ramírez, R. (2016). Colaboración y actividades de innovación en Pymes. // *Contaduría y administración*. 61:3, 568-581. <https://doi.org/10.1016/j.cya.2015.05.016>
- Naciones Unidas. (2022). Informe de los Objetivos del Desarrollo Sostenible 2022. Departamento de Asuntos Económicos y Sociales (DESA).
- Nelson, R. (2008). What enables rapid economic progress: What are the needed institutions? // *Research Policy*. 37:1 (2008), 1-11. <https://doi.org/10.1016/j.respol.2007.10.008>
- Panamá. Asamblea Nacional (2018). Anteproyecto 201 “Para la Recolección y Tratamiento de Desechos Tecnológicos”, de 26 de marzo de 2018. https://asamblea.gob.pa/APPS/SEG_LEGIS/PDF_SEG/PDF_SEG_2010/PDF_SEG_2018/2018_A_201.pdf
- Panamá. Asamblea Nacional (2020). Anteproyecto 164 “Que incentiva y promueve el reciclaje de residuos eléctricos y electrónicos en la República de Panamá y dicta otras disposiciones”, de 30 de septiembre de 2020. https://asamblea.gob.pa/APPS/SEG_LEGIS/PDF_SEG/PDF_SEG_2020/PDF_SEG_2020/2020_A_164.pdf
- Pérez, C. (1996). La Modernización Industrial en América Latina y La Herencia de La Sustitución de Importaciones. // *Comercio Exterior*. 46:5. 347-363.
- Prasad, M. N. V.; Vithanage, M.; Borthakur, A. (Eds.). (2020). *Handbook of electronic waste management: International best practices and case studies*. Butterworth-Heinemann.
- Shittu, O. S.; Williams, I. D.; Shaw, P. J. (2021). Global E-waste management: Can WEEE make a difference? A review of e-waste trends, legislation, contemporary issues and future challenges. // *Waste Management*. 120, 549-563. <https://doi.org/10.1016/j.wasman.2020.10.016>
- Wagner, M.; Baldé, C.; Luda, V.; Nnorom, I.; Kuehr, R.; lattoni, G. (2022). Monitoreo regional de los residuos electrónicos para América Latina: Resultados de los trece países participantes en el proyecto UNIDO-GEF 5554, Bonn (Alemania).
- Zazo, Á.; Ardines, S.; Castro, E. (2015). Redes de colaboración de las unidades de investigación de la Universidad de Panamá: Investigación, desarrollo e innovación. // *Redes: Revista hispana para el análisis de redes sociales*. 26:2, 84-117. <https://raco.cat/index.php/Redes/article/view/303165>

Enviado: 2023-04-10. Segunda versión: 2023-09-15.
Aceptado: 2023-11-02.

Inteligencia artificial y decisiones judiciales: crónica de una transformación anunciada

Artificial Intelligence and judicial decisions: chronicle of an announced transformation

Carolina SANCHIS CRESPO

Universitat de València (España), carolina.sanchis@uv.es

Resumen

La penetración de la IA (inteligencia artificial) en nuestra vida cotidiana es día a día más perceptible. Su presencia va imponiéndose de forma sutil y a la vez imparable. Nuestro sistema de Justicia no podía ser ajeno a esta realidad. Pero mientras hay versiones débiles de la IA que no plantean apenas problemas jurídicos ni éticos y simplemente nos facilitan la vida, la IA fuerte, capaz de tomar decisiones por ella misma, causa recelo y animadversión a partes iguales. Dos preguntas fundamentales se vislumbran desde esa perspectiva. Una, si es posible que en un futuro no muy lejano un juez robot pueda decidir sobre nuestro patrimonio y/o libertad. Otra, si es conveniente y seguro que lo haga. A ambos interrogantes pretende dar respuesta este artículo. Para ello se hace uso de la metodología tradicional en Derecho que consiste básicamente en consultar un número significativo de fuentes bibliográficas y normativa aplicable (aún en ciernes) para, una vez nutrido así el conocimiento sobre la materia, formular nuestro planteamiento particular sobre el tema. No ha sido posible consultar jurisprudencia por razones evidentes. Finaliza el artículo con un apartado dedicado a reflexionar mirando al futuro, en el que se contienen conclusiones y recomendaciones de *lege ferenda* para una próxima regulación de la IA en nuestro ordenamiento jurídico.

Palabras clave: Inteligencia artificial. Decisiones judiciales. Juez robot. Motivación de las decisiones judiciales. Independencia judicial. Imparcialidad judicial.

1. Introducción: la Justicia tecnológicamente transformada y la IA

Tecnología y proceso judicial son dos realidades que históricamente no han ido de la mano. Ha sido lamentablemente habitual que las novedades tecnológicas de las que se ha ido beneficiando la sociedad en el pasado, trascendieran al proceso sólo tras largos años de espera. Finalmente, la Justicia acababa por ser tecnológicamente transformada, aunque siempre con evidente retraso.

Abstract

The penetration of AI (artificial intelligence) in our daily lives is becoming increasingly apparent. Its presence is subtly but unstoppably gaining ground. Our justice system cannot remain unaware of this reality. While there are weak versions of AI that pose hardly any legal or ethical problems by simply making our lives easier, strong AI, capable of making decisions on its own, causes suspicion and animosity in equal proportion. Two fundamental questions stem from this perspective. First, whether it is possible that in a nearby future a robot judge will be able to decide on our wealth and/or freedom. Second if it is convenient and safe for it to do so. This article aims to answer both questions. To do so, we use the traditional methodology in law, consisting in consulting a significant number of bibliographical sources and applicable regulations (still work in progress) to formulate our particular approach to the subject. No jurisprudence was consulted for obvious reasons. The article closes with a reflection looking to the future, containing conclusions and *lege ferenda's* recommendations for a future regulation of AI in our legal system.

Keywords: Artificial Intelligence. Judicial decisions. Robot judge. Judicial decisions motivations. Judicial independence. Judicial impartiality.

Pensemos, por poner un ejemplo lejano, en cuánto tiempo hubo de pasar desde que se comenzó a escribir (al fin y al cabo, la escritura es una clase de tecnología, aunque sea rudimentaria) hasta que la escritura se incorporó al proceso definitivamente con la Decretal de Inocencio III en el año 1216. Y si queremos un ejemplo mucho más reciente, no tenemos más que observar cómo la prueba por documentos electrónicos se incorporó a nuestra Ley de Enjuiciamiento Civil en los inicios del siglo XXI cuando la sociedad llevaba utilizando esta clase de soportes decenas de años.

Con la inteligencia artificial (en adelante IA) como tecnología actual, hubiera podido pasar lo mismo. Sin embargo, sucede que las posibilidades que ofrece a efectos de simplificar el trabajo de los diversos agentes implicados en el proceso (jueces, abogados, letrados de la administración de Justicia, etc.) han hecho que se vaya introduciendo de forma paulatina, inadvertida y por eso mismo, irregular, pues ni siquiera disponemos de una regulación acotada de la misma.

La Justicia tecnológicamente transformada por la IA va a ser distinta a la que conocemos hasta ahora. No nos referimos a la transformación que puede suponer la intervención asistencial de la IA en la tarea del juez. Ahí se tratará de racionalizar y allanar el trabajo cotidiano del juez librándole de tediosas tareas como, por ejemplo, la lectura de miles de folios de un sumario cuando la información relevante esté sólo en unos pocos.

La transformación tecnológica de la Justicia más extrema llegará cuando el juez robot entre en escena. Porque ese momento, nos atrevemos a pronosticar, llegará. No hay más que echar un vistazo a lo que está sucediendo en otros países como China o lo que se prevé para Estonia, cuyos ejemplos detallados omitimos aquí por ser sobradamente conocidos.

Y cuando estemos frente a esa situación será mejor estar preparados y haber repensado lo que sea necesario para adaptarnos a tal realidad. Y hacerlo sin miedo —aunque el miedo a lo desconocido sea una constante en la historia de la humanidad— resultándonos así más cómodo aferrarnos a las verdades tradicionales, alegando que lo que pueda alterarlas mejor que no se incorpore al proceso.

Esa actitud sólo conduce a un negacionismo absurdo que en nada ayuda a comprender cómo debe ser la Justicia tecnológicamente transformada con la que tendremos que lidiar en este siglo. Una Justicia moderna en la que las garantías procesales deben respetarse, aunque haya que hacer cambios para lograrlo.

Si miramos atrás desde mediados del siglo pasado veremos como la IA ha ido escalando posiciones a través del tiempo entreverándose cada vez más en las tareas jurídicas. Desde las bases de datos inteligentes de jurisprudencia, pasando por la Jurimetría y los sistemas expertos [véase al respecto *in extenso* (Barona Vilar, 2021, p. 343-380), hemos llegado en la actualidad a plantearnos como decía *supra*, no sólo la asistencia de la IA al juez humano —algo que casi nadie discute—, sino hasta la posible sustitución de este último por una máquina inteligente.

Esto último es verdaderamente “un salto cualitativo muy significado, que la máquina emule el razonamiento jurídico en la toma de decisiones [...]” (Gómez Colomer, 2021, p. 252). No podemos estar más de acuerdo. Es más, se trata de un auténtico salto en el vacío, pues en la actualidad no disponemos de un verdadero marco normativo por más que éste se demande incluso desde sectores que son conocidos por su falta de apego a la regulación, como aquellos que trabajan en la órbita de Silicon Valley.

Sin ir más lejos, en una comparecencia reciente ante el Senado de los Estados Unidos, Sam Altman, CEO de *Open AI*, solicitó a los gobiernos que adoptasen una regulación que mitigase los crecientes riesgos que plantean los grandes sistemas de IA. Según este directivo (La Ley, 2023, p. 1)

[...]si bien creemos que los beneficios que pueden aportar las herramientas que hemos desplegado hasta ahora, superan con creces los riesgos que plantean, garantizar su seguridad es vital para nuestro trabajo, porque queremos ser claros: si esta tecnología sale mal, puede salir bastante mal

Veamos, para comenzar este trabajo, qué es exactamente la IA y sus aplicaciones en la Justicia, para pasar después a referirnos a las iniciativas regulatorias en torno a la misma. Para ello examinaremos a nivel internacional, y por ser un país puntero en temas tecnológicos donde la IA nació, qué se ha hecho en USA. Seguidamente analizaremos las iniciativas en la Unión Europea que nos vinculan como Estado miembro que somos. Terminaremos este primer epígrafe centrándonos en los objetivos que se pretenden alcanzar con este artículo.

1.1. ¿Qué es la IA y dónde podría aplicarse en la Justicia?

Haciendo un repaso rápido y retrotrayéndonos a los inicios de la IA podemos señalar que en 1943 Mc Culloch y Pitts presentaron un modelo de neurona artificial. Poco después, en 1950, el célebre matemático Alan Turing propuso una prueba para determinar si una máquina podía comportarse de manera inteligente como un ser humano (Morales, 2023, p. 2).

Pero no fue hasta el 31 de agosto de 1955 cuando un grupo de investigadores de Estados Unidos redactó un breve documento en el que solicitaban financiación para un proyecto de investigación de verano para el curso siguiente. El proyecto generó grandes expectativas y sus proponentes predijeron que sería posible crear una máquina tan inteligente como un humano antes de finales de siglo, lo que, como sabemos, finalmente no sucedió.

La propuesta fue firmada por cuatro informáticos muy influyentes: Claude Shannon, el padre de la teoría de la información; Marvin Minsky, uno de los primeros en construir una red neuronal aleatoria; Nathaniel Rochester, el responsable del diseño del IBM 701, el primer ordenador comercial de propósito general; y John McCarthy, al que se atribuye la acuñación del propio término de inteligencia artificial. Esta reunión en Dartmouth se considera el acta fundacional de la IA (Barrio Andrés, 2022, p. 14- 15).

Posteriormente, en 1963, Joseph Weizenbaum creó *Eliza* que fue uno de los primeros programas en procesar lenguaje natural y posibilitó una conversación coherente entre un humano y una máquina. En 1997 *Deep Blue*, computadora de ajedrez creada por IBM, derrotó al campeón mundial de ajedrez Garry Kasparov.

Ya en tiempo más reciente, en concreto en 2011, *Watson*, una computadora de procesamiento de lenguaje natural también de IBM, ganó un famoso concurso de preguntas y respuestas, *Jeopardy!* superando a los concursantes humanos. Posteriormente, en 2016, *AlphaGo*, un programa de IA desarrollado por Google, derrotó al campeón mundial del juego Go, Lee Sedol (Morales, 2023, p. 2).

Mucho más recientemente ha irrumpido a nivel general la tecnología ChatGPT. Se trata de una clase de chatbot de IA desarrollado por la organización de investigación estadounidense OpenAI. Su lanzamiento tuvo lugar en noviembre de 2022 y en 2023 ha sido mundialmente conocida.

En la actualidad es un lugar común decir que (Rivas Velasco, 2023, p. 3)

[...] pese al recurrente uso del vocablo (IA), no existe una definición unánimemente aceptada, ya que mientras que para unos tiene que ver con formas de vida artificial que pueden superar la inteligencia humana, para otros, cualquier tecnología de tratamiento de datos puede considerarse como tal (recomendaciones personalizadas de contenidos de Netflix, Spotify, etc.).

Una buena manera de aproximarnos al concepto de IA es acudir al diccionario de la RAE que la define como una:

Disciplina científica que se ocupa de crear programas informáticos que ejecutan operaciones comparables a las que realiza la mente humana, como el aprendizaje o el razonamiento lógico.

Desde el punto de vista técnico se distinguen fundamentalmente dos clases de IA: la débil y la fuerte. La IA débil (Barrio Andrés, 2022, p. 19)

es la actualmente disponible, tiene la capacidad de resolver problemas específicos, normalmente mejor

que los humanos [...] el objetivo de estos sistemas de IA débil es resolver problemas delimitados.

La IA fuerte (Barrio Andrés, 2022, p. 19)

también conocida como IA general [...] se define como la capacidad de razonar, representar el conocimiento, planificar, aprender, comunicarse en lenguaje natural e integrar todas esas habilidades hacia un objetivo común.

Para entenderlo mejor, un ejemplo de IA débil serían los asistentes de voz Alexa, de Amazon o Siri, de Apple. Responden, comunicándose con nosotros mediante comandos de voz, a la mayoría de preguntas que les formulamos y que tengan que ver con datos. Lo hacen de forma mucho más rápida y eficiente de lo que cualquier ser humano haría.

En cambio, un ejemplo de IA fuerte sería un robot capaz de hallar la solución a un problema de forma análoga a cómo lo haría un ser humano. Debería ser capaz de analizar la situación y hacer todo lo posible con los medios de que dispusiera para lograr su propósito.

Ahora bien, si nos preguntamos dónde podría aplicarse la IA a la Justicia, la respuesta sería muy amplia. Los usos de la IA en nuestro sistema de Justicia pueden ser muchos y variados:

- desde su implementación para automatizar el proceso judicial y consecuentemente la tarea de todos aquellos profesionales que contribuyen a dotar al proceso de la fluidez necesaria (los letrados de la Administración de Justicia, los gestores procesales y administrativos, los tramitadores procesales y administrativos, los auxiliares judiciales, etc.),
- pasando por dotar al ciudadano de un mejor acceso a la Justicia,
- hasta su posible participación en la toma de decisiones judiciales.

Coherentemente con el título de este artículo, nosotros vamos a referirnos a esta última eventualidad.

Así, nos centramos en la figura del juez robot que sería capaz de decidir de forma similar a como un juez humano lo haría. Es decir, teniendo en cuenta cuáles han sido los datos aportados y probados en el proceso y cuáles son los preceptos jurídicos aplicables, debería ser capaz de construir de forma coherente la decisión a tomar.

Los autores favorables a una implantación futura de la IA en la toma de decisiones automatizadas (entre los que me incluyo), coinciden en atribuir determinadas características a los procesos cuyas decisiones podrían ser asumidas por un juez robot. Véase por ejemplo a Delgado Martín

(2022, p. 14). Según este autor, a pesar de que se generan incertidumbres jurídicas, técnicamente es posible la resolución automatizada de un conflicto por parte del sistema de justicia, cuando concurren los siguientes elementos:

1. procesos que se repiten con un contenido idéntico o muy similar (litigación en masa);
2. homogeneidad o gran similitud de las acciones de los demandantes y/o de las contestaciones;
3. decisión jurisdiccional uniforme en la gran mayoría de los casos;
4. escasa complejidad; y
5. presencia de prueba documental.

Como ejemplo de procesos que cumplirían estas características se enumeran los siguientes en el informe del grupo de expertos de trabajo del CGPJ sobre IA (Velasco Núñez, y otros, 2022, págs. 15-19):

- condiciones generales de la contratación
- cláusulas suelo
- cláusulas abusivas
- gastos hipotecarios
- usura TAE tarjetas revolving
- mutuo acuerdo en familia, contrato de transporte aéreo
- cárteles

- alimentos
- conformidades
- juicios rápidos
- proceso por decreto
- asilo
- extranjería
- tráfico
- reclamaciones de intereses
- ejecuciones sin oposición
- monitorios sin oposición
- juicios verbales
- allanamientos de jurisdicción civil
- etc.

Para finalizar este epígrafe vamos a preguntar a la propia IA acerca de su penetración en la Administración de Justicia española. Esto permite consultar las fuentes y comprobar su identidad con la información que proporciona esta herramienta. La autora de este artículo así lo ha hecho. Veremos que nos devuelve la Tabla I de resultados, en la que puede apreciarse que el uso de esta tecnología va a incrementarse exponencialmente en los próximos años. En el primer caso —la estrategia nacional—, hemos preguntado a *perplexity AI* que tiene la ventaja, frente a ChatGPT, de nutrirse también de datos de Internet.

Medidas	Descripción
Estrategia Nacional de IA	España busca potenciar el uso de la IA en la administración pública, incluyendo la Justicia, como parte de su Estrategia Nacional de Inteligencia Artificial
Regulación y marco normativo	Se han realizado avances normativos tanto a nivel nacional como internacional para regular el uso de la IA en la Administración de Justicia, incluyendo el Anteproyecto de Ley de Eficiencia Digital
Desarrollo de capacidades digitales	En el marco del Plan Justicia 2030, se han desarrollado actuaciones para mejorar los servicios digitales en la Administración de Justicia, lo que incluye el uso de tecnologías como la IA.
Implementación de herramientas de IA	El Ministerio de Justicia ha implementado tecnología basada en inteligencia artificial, como la anonimización de documentos y un buscador inteligente, para mejorar sus servicios

Tabla I. Penetración de la IA en la Administración de Justicia española (medidas y descripciones)

Si preguntamos más concretamente por servicios la respuesta es más completa, como se puede apreciar en la Tabla II, en la página siguiente.

1.2. Iniciativas regulatorias en torno a la IA: USA y Unión Europea

Existen en la actualidad diferentes documentos de carácter ético, pero no vinculante, que han ido surgiendo en los últimos años al mismo tiempo

que las potencialidades de la IA iban poniéndose de manifiesto. Sólo algunas de esas iniciativas aspiran a conformar un verdadero marco regulatorio de la IA. Nos referiremos, por razones de espacio, a las más relevantes.

1.2.1. USA

En octubre de 2022 la Casa Blanca presentó una propuesta de Declaración de Derechos ante la IA (Blueprint for an AI Bill of Rights. Making automated systems work for the american people) (1).

<i>Servicio/Iniciativa</i>	<i>Descripción</i>
Anonimización de documentos	Herramienta que permite ocultar o reemplazar caracteres en documentos para reducir riesgos de localizar a personas o utilizar sus datos de manera fraudulenta
Resumen automatizado de documentos judiciales en lenguaje sencillo	Sistema creado para mejorar la comprensión de procesos judiciales, haciendo la justicia más accesible y cercana, promoviendo el acceso equitativo y protegiendo a las personas más vulnerables.
Buscador inteligente de la Justicia (Delfos)	Servicio que ahorra tiempo a los profesionales del sector al permitirles buscar información a través de un lenguaje sencillo
Aplicación de algoritmos para agilizar procesos	Los juristas y el Gobierno aceleran la aplicación de algoritmos para agilizar procesos judiciales, a la espera de la regulación europea
Participación en el Plan de Digitalización de la Administración Pública	El Ministerio de Justicia recibirá parte de los 2.600 millones de euros del Plan de Digitalización de la Administración Pública para implementar soluciones de IA
Jornadas sobre Inteligencia Artificial en la Administración de Justicia	Eventos que analizan las ventajas y reparos que plantea la implantación de soluciones de IA en el ámbito de la Justicia, dando voz a diferentes actores y operadores del ámbito jurídico y tecnológico

Tabla II. Penetración de la IA en la Administración de Justicia española (servicio/iniciativa y descripciones)

La propuesta incluye: un prólogo (2 págs.), cinco principios (7 págs.) y un manual denominado “De los principios a la práctica” (61 págs.), que ofrece medidas concretas que pueden adoptar muchos tipos de organizaciones —gobiernos de cualquier clase y empresas, sin importar su tamaño—, para defender los principios.

Los cinco principios son los siguientes:

1. Protección contra sistemas inseguros y no efectivos.
2. Protección contra la discriminación algorítmica.
3. Protección de la privacidad de los datos.
4. Información del uso del sistema y explicabilidad de sus consecuencias.
5. Garantía de posible intervención humana y retroceso.

El manual *De los principios a la práctica* va repasando cada uno de los cinco principios y aportando modos de hacerlos efectivos en el día a día de gobiernos y empresas. Es particularmente interesante, por su relación con el tema que nos ocupa, el tratamiento que hace del último de los principios: garantía de intervención humana y retroceso.

Según el manual, hay muchas razones por las que las personas prefieren no usar un sistema automatizado. El público estadounidense merece la seguridad de que —cuando los derechos, las oportunidades o el acceso están en juego y existe una expectativa razonable de una alternativa a un sistema automatizado— puedan optar convenientemente por un sistema automatizado y no estén en desventaja por esa elección.

En el sistema de justicia penal (así como en otras áreas sensibles) los sistemas automatizados se usan para muchos propósitos. Sin las apropiadas salvaguardas, estas tecnologías pueden llevar a injustos, imprecisos o peligrosos resultados. Estas áreas sensibles requieren un extra de protección. Es de vital importancia que haya una amplia supervisión humana en tales entornos.

Considerados en su conjunto, los cinco principios y las prácticas asociadas del plan para una *Declaración de Derechos de IA* forman un conjunto de medidas de respaldo contra daños potenciales. Este marco, intencionalmente superpuesto, cuando se toma como un todo, conforma un modelo para ayudar a proteger al público de cualquier daño. Las medidas tomadas para hacer realidad la visión expuesta en este marco debe ser proporcionales a la magnitud y naturaleza del daño o al riesgo de daño, a los derechos, oportunidades y accesos de las personas.

Con un propósito similar a la propuesta de *Declaración de Derechos ante la IA*, el National Institute of Standards and Technology (NIST) presentó un marco para la gestión de los riesgos derivados del uso de la IA en enero de 2023: el *Artificial Intelligence Risk Management Framework*.

A lo largo de sus 48 páginas desgrana toda una batería de medidas para afrontar los riesgos que la IA plantea acorde con su idea de que la gestión de riesgos de la IA es un elemento clave del desarrollo y uso responsable de los sistemas de IA.

Este marco se divide en dos partes. En la segunda se describen cuatro funciones aplicables a la gestión del riesgo aplicables por gobiernos y empresas. Se trata de: la gobernanza, el mapeo, la medición y la gestión.

- La *gobernanza*: Una gobernanza sólida puede impulsar y mejorar las prácticas y normas internas para facilitar la cultura de riesgo de la organización.
 - Las autoridades de gobierno pueden determinar las políticas generales que dirigen la misión, los objetivos, los valores, la cultura y la tolerancia al riesgo de una organización.
 - La alta dirección marca la pauta de la gestión de riesgos en una organización y, con ella, la cultura organizativa. La dirección adapta los aspectos técnicos de la gestión de riesgos de IA a las políticas y operaciones.
 - La documentación puede aumentar la transparencia, mejorar los procesos de revisión humana y reforzar la responsabilidad de los equipos de sistemas de IA.
- El *mapeo*: Esta función establece el contexto para enmarcar los riesgos relacionados con un sistema de IA. El ciclo de vida de la IA consta de muchas actividades interdependientes en las que interviene un conjunto diverso de actores. En la práctica, los agentes de IA encargados de una parte del proceso a menudo no tienen pleno control sobre otras partes y sus contextos asociados. Las interdependencias entre estas actividades y entre los agentes de la IA pueden dificultar una previsión fiable de las repercusiones de los sistemas de IA. Como resultado, las mejores intenciones dentro de una dimensión del ciclo de vida de la IA pueden verse socavadas por interacciones con decisiones y condiciones en otras actividades posteriores.
- La *medición*: La función de medición emplea herramientas, técnicas y metodologías cuantitativas, cualitativas o de metodología mixta

para analizar, evaluar, comparar y supervisar el riesgo de IA y los impactos relacionados. Utiliza el conocimiento que proporciona los riesgos de IA identificados en la función de mapeo e informa la función de gestión. Los sistemas de IA deben probarse antes de su implementación y periódicamente durante su funcionamiento. Las mediciones del riesgo de IA incluyen la documentación de aspectos de funcionalidad y fiabilidad de los sistemas.

- La *gestión*: La función de gestión implica asignar recursos de riesgo a los riesgos mapeados y medidos periódicamente y según lo definido por la función gobernanza. El tratamiento de riesgos comprende planes para responder a incidentes o sucesos, recuperarse de ellos y comunicarlos.

Una vez completada la función de gestión, se dispondrá de planes de priorización de riesgos y de seguimiento y mejora periódicos. Los usuarios del marco dispondrán de una mayor capacidad para gestionar los riesgos de los sistemas de IA implementados y para asignar recursos de gestión de riesgos en función de los riesgos evaluados y priorizados. Corresponde a los usuarios del marco seguir aplicando la función de gestión a los sistemas de IA desplegados a medida que los métodos, contextos, riesgos y necesidades o expectativas de los actores relevantes de la IA evolucionen con el tiempo.

Como vemos estos dos documentos reseñados demuestran la importancia que se da en USA a la gestión de los riesgos que provoca el uso de los sistemas de IA. Acorde con el estilo de este país, las recomendaciones que se dan tienen un marcado carácter práctico que nos parece interesante por su carácter ejemplificador.

1.2.2. Unión Europea

En el seno de la Unión Europea (en adelante UE) ha habido una multitud de iniciativas la última de las cuales es, sin duda, la más significativa:

- Carta ética europea sobre utilización de la IA en los sistemas judiciales y su entorno
- Resolución del Parlamento Europeo, de 6 de octubre, sobre la IA en el Derecho Penal y su utilización por las autoridades policiales y judiciales en asuntos penales
- Borrador cero del Convenio marco sobre IA, Derechos humanos, Democracia y Estado de Derecho
- Ley de IA o Artificial Intelligence Act

Véamoslas con detalle a continuación.

(a) *Carta ética europea sobre utilización de la IA en los sistemas judiciales y su entorno*

En primer lugar, tenemos, en 2018 la *Carta ética europea sobre utilización de la IA en los sistemas judiciales y su entorno*. En ella se proclaman cinco principios inspiradores, a saber:

1. *Principio de respeto de los derechos fundamentales*: asegurar que el diseño y la implementación de las herramientas y servicios de inteligencia artificial sean compatibles con los derechos fundamentales.
2. *Principio de no discriminación*: específicamente se trata de prevenir el desarrollo o la intensificación de cualquier discriminación entre individuos o grupos de individuos
3. *Principio de calidad y seguridad*: con respecto al procesamiento de decisiones judiciales y datos, usar fuentes certificadas y datos intangibles con modelos concebidos en un sistema multidisciplinario en un entorno tecnológico seguro
4. *Principio de transparencia, imparcialidad y equidad*: hacer métodos de procesamiento de datos accesibles y comprensibles que permitan auditorías externas
5. *Principio "bajo control del usuario"*: se trata de no usar un enfoque prescriptivo y asegurar que los usuarios sean actores informados que tengan el control de sus elecciones

Como puede verse, se aprecian similitudes significativas entre estos principios y los que recoge la *Declaración de Derechos ante la IA* estadounidense. En concreto los principios (2), (3), (4) y (5) de la Carta europea encuentran su reflejo, más o menos completo, en los principios (2), (1), (2) y (5) de la Declaración.

(b) *Resolución del Parlamento Europeo, de 6 de octubre, sobre la IA en el Derecho Penal y su utilización por las autoridades policiales y jurisdiccionales en asuntos penales*

Tenemos que trasladarnos al año 2021 para encontrar la *Resolución del Parlamento Europeo, de 6 de octubre, sobre la IA en el Derecho Penal y su utilización por las autoridades policiales y jurisdiccionales en asuntos penales*. Recordemos que esta clase de actos jurídicos no son vinculantes y se limitan a establecer compromisos o posiciones políticas.

- Esta norma considera que toda herramienta de IA desarrollada o utilizada por las autoridades policiales o judiciales debe, como mínimo, ser segura, robusta, fiable y apta para su fina-

lidad, así como respetar los principios de minimización de datos, rendición de cuentas, transparencia, no discriminación y explicabilidad, y que su desarrollo, despliegue y uso deben estar sujetos a una evaluación de riesgos y a una estricta comprobación de los criterios de necesidad y proporcionalidad, debiendo guardar proporción las salvaguardas con los riesgos identificados.

- Destaca que la confianza de los ciudadanos en el uso de la IA desarrollada y utilizada en la Unión, está supeditada al pleno cumplimiento de estos criterios.
- A continuación, subraya la asimetría de poder entre quienes emplean tecnologías de IA y quienes se encuentran sujetos a las mismas.
- Y remarca después que todas las decisiones con efectos legales deben ser tomadas siempre por un ser humano al que puedan pedirse cuentas de las decisiones adoptadas.
- Considera que todas las personas objeto de sistemas de IA deben poder acceder a vías de recurso y recuerda que, en virtud del Derecho de la Unión, una persona tiene derecho a no ser objeto de una decisión que produzca efectos jurídicos que la conciernen o la afecte significativamente y que se base únicamente en el tratamiento automatizado de datos.

(c) *Borrador cero del Convenio marco sobre IA, Derechos humanos, Democracia y Estado de Derecho*

Posteriormente, el 6 de enero de 2023, el Comité sobre IA del Consejo de Europa publicó un borrador cero del Convenio marco sobre IA, Derechos humanos, Democracia y Estado de Derecho. En él se recogen principios a los que ya nos hemos referido en documentos anteriores, como el principio de seguridad o el de transparencia y supervisión. Se prevén también mecanismos de gestión de riesgos e impactos causados por sistemas de IA. Este documento será negociado con la Comisión Europea en el futuro y vinculará a las Partes firmantes que deberán adoptar las medidas necesarias en su legislación nacional para dar efecto a los principios, normas y derechos enunciados en el futuro Convenio. Con el fin de garantizar la aplicación efectiva de sus disposiciones por las Partes, el Convenio establecerá un mecanismo de seguimiento.

(d) *Ley de IA o Artificial Intelligence Act*

Pero sin duda, la fecha clave en materia de normativa sobre IA a nivel europeo, es la del pasado 14 de junio de 2023. Ese día el Parlamento Europeo adoptó una posición común sobre el texto

final del Reglamento de IA, la que se ha dado en llamar *Ley de IA* o *Artificial Intelligence Act* por su nombre en inglés. Será previsiblemente la primera regulación de la IA una vez concluya la fase de negociaciones entre el Consejo, la Comisión y el Parlamento y vea la luz esta ley pionera a escala mundial. Probablemente a finales de 2023 esté listo el texto definitivo.

El propósito del Reglamento es fomentar la adopción de la inteligencia artificial centrada en el ser humano y confiable. También se trata de garantizar un alto nivel de protección de la salud, la seguridad, los derechos fundamentales, la democracia y el Estado de Derecho y el medio ambiente, frente a los efectos nocivos de los sistemas de inteligencia artificial en la Unión, apoyando al mismo tiempo la innovación y mejorando el funcionamiento del mercado interior” (2).

El concepto de IA que se maneja se identifica con una tecnología en constante evolución que puede contribuir, y de hecho ya contribuye, a una amplia gama de beneficios económicos, medioambientales y sociales en todo el espectro de industrias y actividades sociales si se desarrolla de acuerdo con los principios generales pertinentes, en consonancia con la Carta y los valores en los que se fundamenta la Unión (3).

Una cuestión discutida en la que finalmente se ha llegado a un acuerdo es la definición de los sistemas de IA prohibidos. Entre ellos se encuentran:

los sistemas de identificación biométrica remota en tiempo real en espacios de acceso público; los sistemas de identificación biométrica a posteriori (con una excepción a favor de la policía y tras la pertinente autorización judicial); y los sistemas de reconocimiento de emociones.

Especialmente relevante para el tema de este trabajo es lo que se dispone en la enmienda 71 en la que se clasifican las decisiones automatizadas en justicia como de alto riesgo habida cuenta de su impacto potencialmente significativo en la democracia, el Estado de Derecho, las libertades individuales y el derecho a la tutela judicial efectiva y a un juicio justo.

En particular, para hacer frente a los riesgos de posibles sesgos, errores y opacidad, es conveniente calificar de alto riesgo los sistemas de IA destinados a ser utilizados por una autoridad judicial o un órgano administrativo o en su nombre para asistir a las autoridades judiciales o a los órganos administrativos en la investigación e interpretación de los hechos y del derecho y en la aplicación del derecho a un conjunto concreto de hechos, o utilizados de forma similar en la resolución alternativa de litigios.

A continuación, se presenta una tabla comparativa de las principales novedades de la AI Act europea frente a la normativa anterior hoy vigente.

Aspecto	Artificial Intelligence Act	Normativa Anterior (vigente a 24 de noviembre de 2023)
Enfoque	Basado en el riesgo, imponiendo obligaciones a los sistemas de IA de alto riesgo	Críticas de que promueve un entorno de IA desregulado en Europa y tiene poco sentido e impacto
Protección de derechos humanos	Pretende garantizar que los sistemas de IA no causen daño a los seres humanos y preservar los derechos fundamentales	Críticas de que los derechos humanos están relegados a un segundo plano y que el documento va en una dirección mercantilista
Prohibición de tecnologías abusivas	Existe preocupación de que la AI Act pueda abrir la puerta al uso de tecnologías que entren en conflicto con el derecho internacional de los derechos humanos	Críticas de que se permiten sistemas de IA nocivos, y que no se prohíben sistemas de elaboración de perfiles racistas y discriminatorios

Tabla III. principales novedades de la AI Act europea frente a la normativa anterior (Perplexity AI)

Estas comparaciones reflejan las diferencias en el enfoque, la protección de los derechos humanos y la regulación de tecnologías abusivas.

Las fuentes bibliográficas utilizadas por la herramienta han sido comprobadas por la autora del artículo.

El uso de herramientas de inteligencia artificial puede apoyar, pero no debe sustituir, el poder de decisión de los jueces o la independencia judicial, ya que la toma de decisiones final debe seguir siendo una actividad y una decisión humana.

1.3. Objetivos de este artículo

Los objetivos de este artículo se centran en dar respuesta a estas dos preguntas:

- una, si es posible que en un futuro no muy lejano un juez robot pueda decidir sobre nuestro patrimonio y/o libertad;
- otra, si es conveniente y seguro que lo haga.

Nos ha parecido pertinente comenzar con una introducción para contextualizar la IA en la Justicia en USA y en la UE, para pasar a continuación a considerar la posibilidad y después la conveniencia de introducir la figura del juez robot en nuestro ordenamiento jurídico. Porque lejos de ser una mera entelequia es algo comúnmente admitido incluso entre aquellos autores más renuentes a aceptar el juicio robótico. En este sentido son clarísimas las palabras de Ignatieff (2023):

Pronto llegará el día en que las máquinas se conviertan en mejores jueces que nosotros [...].

2. Aterrizando el concepto de juez robot

A los efectos de esbozar el concepto de juez robot en nuestro sistema de Justicia, conviene empezar hablando de datos y algoritmos. Se trata de dos realidades intrínsecamente unidas de modo tal que sin datos los algoritmos no podrían ejecutar la IA y sin algoritmos la mera acumulación de datos no sería significativa. Los algoritmos convierten las masas de datos en información relevante. Los algoritmos dan sentido a los datos.

2.1. Algoritmos

Los algoritmos son, según la RAE, un conjunto ordenado y finito de operaciones que permite hallar la solución de un problema. Este conjunto de operaciones se presenta en forma de software cuando hablamos de IA. El material que permite que el algoritmo nos dé resultados no es otro que los datos.

2.1.1. Datos

Tratándose de datos importa mucho la calidad de los mismos pues son la materia prima de la que se nutrirán los algoritmos. Garantizada esa calidad, se ha de disponer de una cantidad significativa de ellos para que las conclusiones que se extraigan sean realmente relevantes. Como señala Delgado Martín (2020, p. 524):

La aplicación de instrumentos de IA exige los siguientes presupuestos; en primer lugar, la existencia de datos con calidad; y, en segundo lugar, un volumen suficiente de datos que pueden proceder de fuentes heterogéneas [...]. En definitiva, cuanto

mayor número de datos, mayores y mejores son las posibilidades de tratamiento.

Sin embargo, nos topamos con un grave problema cuando intentamos hacer minería de datos en el sector de la Justicia, pues como indica el mismo autor (Delgado Martín, 2020, p. 562):

El sistema judicial, al igual que ocurre con otros sectores públicos, no se encuentra configurado para manejar y aprovechar un gran volumen y variedad de datos; y tiene una comprensión de sus activos de datos [...] que puede calificarse de rudimentaria; incluso no es fácil responder a preguntas básicas tales como cuántas bases de datos existen dentro de la organización judicial, qué información concreta contiene cada base de datos, o cómo los datos son recopilados.

Una de las causas de esta desorganización es, sin duda, la existencia de Comunidades Autónomas con competencia de Justicia transferidas (la mayoría), junto a otras dependientes del Ministerio de Justicia. A la hora de informatizar los datos, parece ser que nadie reparó en la necesaria estandarización, y ahora —cuando el poder acceder a datos en calidad y cantidad es un presupuesto imprescindible para poder aplicar las soluciones de IA al sector Justicia— se levanta frente a esta tarea un muro aparentemente infranqueable.

Será necesario que a partir de ahora los datos se generen teniendo en cuenta todo lo anterior. En este sentido es remarcable el esfuerzo realizado por el *Manifiesto por un espacio público de datos en el ámbito de Justicia* (Comité Técnico Estatal de la Administración Judicial Electrónica, 2022) en el que se establece el compromiso con el cumplimiento de estándares de calidad en la generación de datos. Se sujeta la producción de datos a los principios de pertinencia, precisión, actualidad, puntualidad, accesibilidad, comparabilidad y coherencia.

La utilización de herramientas de validación y de control de calidad de los datos será promovida desde el sector público. En este sentido, la producción de los datos debe ir orientada a su reutilización ofreciendo datos dinámicos en tiempo real con el mayor nivel de desagregación posible.

El problema sigue siendo qué hacemos con los datos anteriores. Tendremos que asumir su carácter fragmentario e incompleto y con esos mimbres hacer el cesto de la IA cómo mejor se pueda.

2.1.2. Sesgos y ruido

A los algoritmos de IA se les acusa de ser sesgados y a menudo el sesgo se confunde con el ruido cuando conceptualmente son realidades próximas, pero no equivalentes. De hecho, como tendremos oportunidad de ver a continuación, los al-

goritmos pueden ser sesgados, pero no son ruidosos. Sin embargo, las decisiones de las personas pueden estar contaminadas tanto por el sesgo como por el ruido.

Los sesgos son similares a los prejuicios. Cuando se imputa a los algoritmos de IA un sesgo, lo que se quiere decir es que los datos que arroja no son neutrales. En igualdad de circunstancias, el pertenecer a determinado grupo social, género, religión, etc., hace que la solución sea distinta. No garantizan un trato igual entre los diferentes grupos (por ejemplo, si eres de raza negra vas a prisión provisional, si eres de raza blanca a libertad provisional).

Desde este punto de vista está claro que no sólo los algoritmos de IA tienen sesgos, las personas también los tienen. De hecho, el sesgo de los algoritmos proviene originariamente de los sesgos personales que posteriormente se incorporan a los datos. Y como se ha dicho “El sesgo más grande es creer que no tenemos ningún prejuicio” (Peiró & Baeza-Yates, 2022, p. 3).

En nuestra opinión, los sesgos humanos pueden a su vez ser clasificados como voluntarios o involuntarios. Los voluntarios son aquellos que dependen directamente de la voluntad de la persona que lo aplica (por ejemplo, creer que una determinada raza es inferior a otra y por ello juzgarla más severamente en el caso concreto). Los involuntarios actúan con igual eficacia, aunque su origen es distinto. Son circunstancias variadas que impactan en la decisión, pero no entran en la categoría de prejuicio. Estas circunstancias afectan al juez y modifican su decisión inconscientemente.

Las circunstancias pueden ser:

- personales (por ejemplo, el estado de salud del juez);
- ambientales (por ejemplo, la temperatura de la sala de vistas); u
- horarias (por ejemplo, el tiempo transcurrido desde la última ingesta de comida del juez).

Todas esas circunstancias pueden influir en la decisión que finalmente el juez adopte.

El ruido es diferente. No se refiere a un juez en particular sino al sistema en general. Señala la incoherencia de las decisiones judiciales explicando por qué en casos análogos las decisiones de uno u otro juez pueden ser diametralmente opuestas. Los algoritmos están libres de ruido porque, a diferencia de los seres humanos, una fórmula siempre devolverá el mismo resultado cuando los datos que se le suministren sean los mismos.

2.1.3. Mitigación de sesgos: transparencia y evaluaciones de impacto

Visto lo anterior ya sabemos a ciencia cierta que a los algoritmos de IA puede imputárseles, sin miedo a equivocarnos, ser sesgados, aunque no ruidosos. Las preguntas que surgen a continuación son: ¿debemos conformarnos con algoritmos sesgados? ¿no se puede hacer nada para evitarlo? Y la respuesta es afirmativa, sí puede hacerse algo. Podemos mitigar esos sesgos cosa que, nótese, no podríamos hacer con un sesgo humano, al menos a priori (4).

La *mitigación de los sesgos* se hace, fundamentalmente, mediante la aplicación de la transparencia y de las evaluaciones de impacto. Veamos qué significan una y otras.

La transparencia de los algoritmos es una característica de los mismos generalmente exigida en los documentos que establecen estándares éticos y en las normativas en ciernes. Como ejemplo tenemos lo que se establece en la *Artificial Intelligence Act* al respecto (5):

Los sistemas de alto riesgo de IA se diseñarán y desarrollarán de forma que su funcionamiento sea lo suficientemente transparente como para permitir a los proveedores y usuarios comprender razonablemente el funcionamiento del sistema. Se garantizará una transparencia adecuada de acuerdo con la finalidad prevista del sistema, orientada a cumplir las obligaciones pertinentes del proveedor y del usuario. Por transparencia se entenderá que, en el momento en que se comercialice el sistema de alto riesgo de IA, se utilicen todos los medios técnicos disponibles de acuerdo con el estado de la técnica generalmente reconocido, para garantizar que los resultados del sistema de inteligencia artificial puedan ser interpretados por el proveedor y el usuario. El usuario deberá estar capacitado para comprender y utilizar adecuadamente el sistema de IA mediante el conocimiento general de cómo funciona el sistema de IA y qué datos procesa, lo que le permitirá explicar las decisiones tomadas por el sistema de IA a la persona afectada [...].

En nuestra opinión queda desdibujada en la norma transcrita la esencia de la transparencia porque para que el algoritmo de IA sea plenamente diáfano el código fuente debería ser público. Si no lo es, difícilmente podrá saberse “qué factores utiliza, a cuáles se les da más peso o cómo se interrelacionan” (Martínez Garay & García Ortiz, 2022, p. 163). La norma se refiere a “la comprensión razonable” y al “conocimiento general” del funcionamiento del sistema, y eso no incluye el conocimiento detallado que proporciona el código fuente.

Como ejemplo positivo de transparencia podemos citar el art. 16.1 apartado I) de la Ley 1/2022,

de 13 de abril, de Transparencia y Buen Gobierno de la Comunitat Valenciana que establece que:

Las administraciones públicas deben publicar [...]: La relación de sistemas algorítmicos o de inteligencia artificial que tengan impacto en los procedimientos administrativos o la prestación de los servicios públicos con la descripción de manera comprensible de su diseño y funcionamiento, el nivel de riesgo que implican y el punto de contacto al que poder dirigirse en cada caso, de acuerdo con los principios de transparencia y explicabilidad.

Como conclusión cabe decir que sí es posible una mitigación de los sesgos algorítmicos mediante el requisito de transparencia, pero sólo garantizando la publicación de los códigos fuentes en abierto. Mientras los derechos a la propiedad intelectual y al secreto empresarial primen más que la transparencia, ésta no estará asegurada.

En cuanto a las *evaluaciones de impacto* de un sistema de IA, con ese término se hace referencia a los procesos que se llevan a cabo para determinar las consecuencias nocivas futuras de la implementación del mismo. Veamos cómo se contemplaban estos procedimientos en el seno de la UE.

Fueron originalmente identificadas por (Dalli, 2021) (6) y podían producirse en diferentes niveles. En concreto, la evaluación de impacto identificaba seis problemas desencadenados por el desarrollo y el uso de la IA. Serían los siguientes:

1. El uso de la IA plantea mayores riesgos para la seguridad de los ciudadanos, que no están suficientemente cubiertos por el actual marco de seguridad de la UE.
2. El uso de la IA plantea un mayor riesgo de violación de los derechos fundamentales de los ciudadanos y de los valores de la Unión.
3. Las autoridades carecen de las competencias, de marcos de procedimiento y de los recursos necesarios para garantizar y supervisar el cumplimiento de las normas aplicables al desarrollo de la IA.
4. La inseguridad jurídica y la complejidad de la aplicación de las normas existentes en los sistemas de IA disuaden a las empresas de desarrollar y utilizar sistemas de IA.
5. La desconfianza en la IA podría frenar su desarrollo en Europa y reducir la competitividad global de la economía de la UE.
6. Las medidas fragmentadas crean obstáculos para un mercado único transfronterizo de IA y amenazan la soberanía digital de la Unión.

La normativa en desarrollo de la Artificial Intelligence Act reconoce estas vulnerabilidades y propone, antes de poner en funcionamiento un sistema de alto riesgo de IA, una evaluación de impacto sobre los derechos fundamentales que obliga a los responsables del despliegue a examinar, como mínimo, los siguientes elementos:

1. una descripción clara de la finalidad prevista para la que se utilizará el sistema,
2. una descripción clara del ámbito geográfico y temporal previsto para la utilización del sistema,
3. las categorías de personas físicas y grupos que puedan verse afectados por la utilización del sistema,
4. la verificación de que la utilización del sistema se ajusta a la legislación nacional y de la Unión pertinente en materia de derechos fundamentales,
5. el impacto razonablemente previsible del sistema en el contexto específico de su utilización,
6. los riesgos específicos del perjuicio que pueda afectar a las personas marginadas o a los grupos vulnerables,
7. el impacto negativo razonablemente previsible de la utilización del sistema sobre el medio ambiente,
8. un plan detallado sobre cómo se atenuarán los perjuicios y el impacto negativo sobre los derechos fundamentales identificados, y
9. el sistema de gobernanza que establecerá el implementador, incluida la supervisión humana, la tramitación de reclamaciones y la reparación.

Si no puede identificarse un plan detallado para mitigar los riesgos señalados en el curso de la evaluación, el implementador se abstendrá de poner en funcionamiento el sistema de IA de alto riesgo e informará de ello sin demora al proveedor y a la autoridad nacional de supervisión (7).

Así como respecto a la transparencia albergamos serias dudas expresadas supra de que tal y como está concebida en el texto actual de la Artificial Intelligence Act pueda efectivamente mitigar los sesgos del sistema de IA, las evaluaciones de impacto nos merecen un juicio positivo. Conjuramos los riesgos de los que puede adolecer el sistema de IA evaluado y lo hacen a través de un análisis pormenorizado y riguroso de sus posibles vulnerabilidades.

2.2. La función jurisdiccional: ¿está capacitado el juez robot para juzgar?

En la doctrina se perfilan claramente dos sectores respecto a la posibilidad de incorporar en nuestro sistema de Justicia la figura del juez robot capacitado para juzgar de modo análogo a como lo hacen los jueces humanos.

Encontramos, por un lado, a aquellos autores que ni siquiera consideran tal posibilidad (Martín Pastor, 2023, p. 626- 627) para quien

[...] el escenario de los jueces robot presenta dos limitaciones: 1ª) Requeriría una reforma constitucional (arts. 117.1 y 3 CE). 2ª) El estado actual de la tecnología no lo permite.

Y también están los que reconociendo que será una realidad, se decantan por cerrar los ojos ante la misma en un ejercicio de avestruzismo (Gómez Colomer, 2023, p. 325)

[...] en las circunstancias actuales la existencia de un juez-robot sentenciador no debería ser admisible en nuestra organización judicial, ni para nuestro proceso civil, ni para nuestro proceso penal. No vale la pena [...] No estoy contra la IA, sólo estoy en contra de la justicia robótica decisoria. No me convence ni me gusta y no le veo además ninguna ventaja, y sí veo en cambio muchos inconvenientes.

Otros autores (Del Moral García, 2023, p. 6) defienden que

[...] en la tarea estricta de enjuiciar, precisar qué es lo justo, lo de cada uno, —y eso es el núcleo duro de la función judicial— el factor humano es irrenunciable e imprescindible.

Entendiendo lo que se quiere decir con esto, pensamos que en un sistema de derecho positivo lo que es justo lo define la ley. Nos hemos puesto de acuerdo en que nuestros valores, como sociedad democrática, se plasman en la ley. Por tanto, la función de juzgar consiste en aplicar la ley al caso concreto.

Es cierto que la ley, a veces, proporciona márgenes de discrecionalidad al juez. Nada impide que una máquina pueda ser alimentada con datos que le permitan adecuar esa discrecionalidad al caso concreto y hacerlo con criterios humanos que estarán contemplados en los algoritmos.

Por otro lado, están quienes piensan que el juez robot será una realidad viable, tesis en la que como decíamos supra, nos inscribimos. Así (Martínez Zorrilla, 2019, p. 9) para quien

[...] los avances en inteligencia artificial son rápidos y significativos, y [...] cabe esperar que sus desarrollos aun lo sean en mayor medida en un futuro, con lo que no parece en absoluto descartable la idea de un juez artificial completamente autónomo.

o (Barona Vilar, *ibídem*, 2021) que asegura que

La función jurisdiccional indudablemente puede desempeñarse por una máquina, no por cualquier máquina, sino aquella que permite su conformación computacional algorítmica y se va perfeccionando, cada vez más, hacia los modelos de computación cuántica, y no binaria.

Desde este segundo planteamiento, lo que debe hacerse es esbozar la figura del juez robot aprovechando sus innegables ventajas y sin pérdida de garantías procesales para el justiciable.

Habrà que examinar, pues, si el juez robot será capaz de determinar los hechos controvertidos del caso, si sabrà determinar el derecho aplicable y si estará preparado para ofrecer una motivación suficientemente reveladora de cuáles fueron las razones que le condujeron a esa decisión. Es decir, se trata de saber si el juez robot será capaz de cumplir la función jurisdiccional.

2.2.1. Determinación de los hechos

Con el estado actual de la tecnología es indudable que se puede crear un instrumento de IA en el que puedan introducirse los elementos probatorios de un proceso de modo tal que se pueda extraer el resultado de cuál sería la hipótesis más plausible de hechos probados. También podría informar acerca de la suficiencia probatoria de los datos introducidos.

De esta manera funcionan hoy en día las soluciones de IA que juzgan en China o las que previsiblemente lo harán en Estonia, aunque su desarrollo hoy no es una realidad. Sin embargo, no podemos desconocer que, en el estado actual de la tecnología, no estamos todavía ante la posibilidad de que el uso de esta herramienta sea autónomo y sigue siendo necesaria la supervisión por un juez humano.

Como ejemplos de estas herramientas refiere Nieva Fenoll (2018, p. 26-27), en su obra pionera en el tratamiento de la IA en el proceso judicial español, los siguientes:

- *Stevie*, programa que construye historias coherentes partiendo de los datos existentes; *Echo* y *Peirce-IGTT*, aplicación que elabora hipótesis y estrategias de acusación y defensa; y
- el programa *Alibi*, que ante un determinado delito hace un pronóstico de las diferentes explicaciones que pueda tener el comportamiento del reo (*sic*) a los efectos de comprobar esas explicaciones, incluso aunque el reo (*sic*) decidiera no ofrecerlas.

La principal ventaja de que sea un juez robot y no un juez humano el que determine la veracidad de

los datos aportados al proceso y llegue así a aislar cuáles son los datos necesitados de prueba, es la objetividad en la selección. Ya hemos visto que los sesgos del algoritmo pueden ser mitigados fundamentalmente con las evaluaciones de impacto. Así, podemos decir que, si esa evaluación de impacto ha sido correctamente aplicada y valida el funcionamiento de la herramienta, el resultado que arroje el algoritmo será confiable.

No cabe duda de que será mejor al principio comenzar aplicando estas soluciones de IA al proceso civil y, dentro de él, en aquellos procesos que, o bien no necesitan pruebas, o bien en los que la dificultad de la prueba es mínima por tratarse exclusivamente de prueba documental. Pero eso no significa que en un futuro cercano no puedan implementarse herramientas que permitan incorporar al proceso todo tipo de pruebas mediante IA.

Pensemos, por ejemplo, en las decisiones que debe tomar un juez humano para aceptar o no la veracidad de las declaraciones de un testigo. El juez echará mano de su experiencia y así juzgará más creíble al testigo que parezca más tranquilo, que sea más preciso en su testimonio, que menos se contradiga, etc.

Esas variables, y otras similares, pueden ser apreciadas con precisión por un software debidamente entrenado: la tranquilidad del testigo puede medirse por los latidos de su corazón y por la ausencia de sudoración; la precisión en el testimonio mediante el análisis del lenguaje natural; las contradicciones de igual manera, etc.

Como síntesis, cabe decir que sí es posible que el software que conocemos como juez robot sea capaz de hacer una selección objetiva de los datos relevantes que conducen a la determinación de los hechos controvertidos en un proceso. En la actualidad, sin embargo, el estado de la tecnología no permite aplicar indiscriminadamente la herramienta a cualquier tipo de proceso ni hacerlo sin supervisión del juez humano.

2.2.2. Determinación del derecho

En lo relativo a la determinación del derecho el tema resulta más escabroso.

Hemos de empezar reconociendo que la aplicación informática que permita introducir los datos normativos aplicables a un caso se encontrará, sin duda, con esta típica limitación a la que alude De la Oliva Santos (2022, p. 32):

[...] es muy probable que la aplicación informática (defectos intrínsecos aparte, aunque suelen ser habituales, por insuficiente formación de los programadores) prescinda de matices o elementos normativos relevantes, que requerirían interpretación. Pero

como no puede consignarse más que lo que la aplicación informática prevea, lo demás no cuenta.

Por otro lado, el viejo aforismo romano *in claris non fit interpretatio* fue desechado hace tiempo pues se reconoció que siempre —por más que la norma a aplicar sea sencilla— será necesario un trabajo de interpretación cuando no de elaboración del derecho. De hecho, ante la norma a aplicar el juez puede encontrarse en una de estas tres situaciones (Sanchis Crespo, 2022, p. 83-84): la inexistencia de la norma, la indeterminación de la norma y la existencia de la norma.

(a) La inexistencia de la norma

En primer lugar, podemos encontrarnos con que el legislador incumpla su misión y deje sin regular situaciones jurídicas. Quizás estas situaciones no existían en el pasado, pero el correr de los años ha hecho que sean cotidianas y mientras el legislador no reaccione nos encontraremos ante un vacío normativo. Encontramos ejemplos en el proceso penal, en el que hasta finales de 2015 no se prescribía cómo debían investigarse los dispositivos de almacenamiento masivo de información como disquetes, DVDs, etc.

¿Qué hacer ante este panorama? Los jueces humanos investigaban estas fuentes de prueba aplicando principios generales del derecho y analogía.

¿Cómo debería comportarse un juez robot? Exactamente igual. El algoritmo debería ser alimentado con los datos suficientes para saber qué hacer si se encuentra en esta situación de anomia. No desconocemos la dificultad de este supuesto. No es lo mismo introducir datos sobre una normativa existente que hacerlo sin ese soporte.

Por este motivo estimamos que será difícil conseguir que el juez robot ejecute esta tarea para la que hace falta finura jurídica difícilmente trasladable a la aplicación informática. No decimos que sea imposible pero sí que este es el supuesto más complicado de resolver por la tecnología de IA. Se trataría ni más ni menos de que el juez robot elaborase judicialmente el derecho.

(b) La indeterminación de la norma

En este otro escenario nos encontramos, no ante la ausencia de norma, pero sí ante la ausencia de norma lo suficientemente concreta. Se trata de las conocidas como normas abiertas o indeterminadas que exigen también elaboración judicial del derecho.

Por ejemplo, el artículo 1104 del Código Civil (CC) establece que

La culpa o negligencia del deudor consiste en la omisión de aquella diligencia que exija la naturaleza de la obligación y corresponda a las circunstancias de las personas, del tiempo y del lugar. Cuando la obligación no exprese la diligencia que ha de prestarse en su cumplimiento, se exigirá la que correspondería a un buen padre de familia.

Como vemos, se dispone como parámetro de la diligencia a exigir cuando la norma no la prevea, la que correspondería a un buen padre de familia, que se toma como tipo medio de persona aceptablemente diligente. Dejando a un lado la formulación machista del precepto que es deudora de otra época, resulta claro que la propia ley establece un patrón de conducta que deberá integrar el juez en cada caso que se someta a su consideración y en el que no haya referencia a la diligencia exigible.

Podemos claramente apreciar las dificultades que surgirán también aquí para que el juez robot pueda autónomamente resolver. En primer lugar, tendrá que apreciar la inexistencia de diligencia exigible y, a continuación, identificar si esa diligencia del buen padre de familia es atribuible al supuesto concreto.

También aquí pensamos, como en el caso anterior, que es preciso desplegar una finura jurídica difícil de conseguir por la herramienta de IA. Quizás se consiga en un futuro, pero de momento esa posibilidad resulta remota.

(c) La existencia de la norma

Finalmente nos encontramos con los casos en los que la norma existe y es completa. Aquí cabría objetar que no hay elaboración judicial y que, por tanto, es más sencillo alimentar al juez robot con esos datos normativos para que de manera automática nos proporcione la decisión judicial adecuada. Sin embargo, no podemos desconocer que la ley está concebida para una generalidad de supuestos y al aplicarla al caso concreto siempre será necesaria una cierta labor de adaptación.

Resumiendo lo expuesto hasta aquí, podemos decir que siempre hay una elaboración judicial del derecho. En unos casos será más evidente (inexistencia de la norma y normas indeterminadas) y en otros menos (existencia de la norma). Hasta qué punto un juez robot pueda llegar a desempeñar esa tarea es algo que, si bien en este momento es un reto a alcanzar, en un futuro próximo creemos que será una realidad.

2.2.3. Decisión judicial: la motivación

El artículo 120.3 de la Constitución Española (CE) establece que “Las sentencias serán siempre motivadas [...]”. Exigencia que las leyes de

enjuiciamiento extienden a otras resoluciones judiciales, en concreto a los autos y a algunas providencias.

Motivar consiste en exteriorizar las razones que conducen a la decisión que se toma. En el caso de una sentencia serán las razones que conducen al fallo.

Siendo pues ésta una exigencia ineludible de las sentencias, debemos preguntarnos si esa motivación puede ser elaborada por un juez robot.

Hay autores que se pronuncian abiertamente en contra de ello (Gómez Colomer, 2023, p. 144):

La máquina no puede razonar, no puede argumentar, no puede explicar por qué decide absolver o condenar, simplemente absuelve o condena, una única palabra sale de su boca digital. Evidentemente una decisión de un juez robot no es ni puede ser una decisión fundada, motivada. Toda la doctrina está de acuerdo en afirmar que nunca llegaremos a tanto en el proceso penal (tampoco en el civil en realidad), al menos, nunca podrá una máquina sustituir a una persona juez cuando se tenga que tomar una decisión que afecte a la libertad de una persona [...].

Frente a esta encendida argumentación en contra de cualquier motivación atribuible a sistemas de IA, hay que comenzar puntualizando que no toda la doctrina está de acuerdo en esas aseveraciones. Por ejemplo, el magistrado de la Audiencia Nacional, Velasco Núñez (2023, p. 4), especialista en IA y miembro del comité de expertos del Consejo General del Poder Judicial (CGPJ) sobre IA, se posiciona a favor del juez robot, como ya ocurre en otros países

[...] porque en tareas repetitivas, de escasa dificultad técnica y mínima aplicación normativa [...] y sobre cuestiones de mínima cuantía, es lo que ya deberíamos estar aplicando para evitar el secular retraso al resolver los problemas de los justiciables y así conseguir que los jueces dediquen su tiempo a las cuestiones más novedosas y difíciles.

Asimismo, Martínez Zorrilla (2019, p. 2 y 3) se pregunta por la posibilidad del que él denomina juez artificial y se responde diciendo que:

Si bien podemos decir que en la actualidad estaríamos aún bastante lejos de contar con procesos integrales artificiales de aplicación del derecho [...] los avances han sido muy significativos en los últimos años [...] con resultados en no pocos casos sorprendentes, que hacen que legítimamente podamos plantearnos la posibilidad de que en el futuro los procesos de aplicación del derecho puedan ser (casi) totalmente automáticos.

Conviene también señalar que, *mutatis mutandi*, pasa con la admisión del juez robot en nuestro ordenamiento y sus funciones (entre las que se

encuentra, sin duda, la motivación de sus decisiones), lo mismo que pasaba a finales del siglo pasado con la admisión de los soportes informáticos como documentos a efectos de prueba en el proceso.

Una parte importante de la doctrina no consideraba que pudieran ser tenidos por tales. El paso de los años demostró que sí eran documentos de pleno derecho y que la firma digital —cumpliendo determinados requisitos— no era menos segura que la tradicional firma manuscrita.

Las sucesivas reformas de la Ley de Enjuiciamiento Civil (LEC) incorporando a los soportes informáticos como medio de prueba documental, terminaron de poner las cosas en su sitio. Quienes lo vislumbramos antes de que se reconociera expresamente por las leyes (Sanchis Crespo, 1999, p. 47-98), tenemos la satisfacción personal de habernos adelantado a los acontecimientos con una intuición que se demostró certera. Con el juez robot y su capacidad de motivación, aventuro que pasará lo mismo. La realidad es testaruda y nuestro ordenamiento jurídico acabará rindiéndose a esa evidencia.

Contrario también a la posibilidad de que los sistemas de IA motiven se muestra Marchena Gómez (2022, p. 54-55):

El derecho a la motivación de las resoluciones jurisdiccionales [...] no puede satisfacerse con una motivación automatizada en la que el algoritmo sustituye al razonamiento judicial. No todo lo que ha de ser objeto de motivación puede ser asumido por mecanismos automatizados.

En la última frase transcrita está la clave, estimamos, para deslindar dónde cabe y dónde no IA en la motivación de las decisiones judiciales hoy por hoy. Habrá que distinguir entre las motivaciones automatizables de aquellas que no lo sean. Es decir, habría que diferenciar entre las motivaciones complejas y las motivaciones más sencillas y reiterativas.

Se trata, en este último caso, de lo que nuestro Tribunal Constitucional (TC) —STC 8/2002, de 14 de enero— da en llamar “resoluciones de modelos impresos o estereotipados” y que vulgarmente se conocen como motivaciones de “corta y pega” en alusión a las funciones del procesador de texto que permiten construir una motivación casi instantánea, haciendo uso de las mismas. Según el alto tribunal

[...] aunque desaconsejable su utilización por ser potencialmente contrarias al derecho a la tutela judicial efectiva, no implica necesariamente una falta o insuficiencia de motivación [...] debiendo analizarse caso por caso para determinar la suficiencia de la respuesta ofrecida [...].

Sintetizando, en los casos de resoluciones estereotipadas aportar una motivación de IA sería tan sencillo como lo es para los jueces humanos recurrir a ellas. No creo que desde este punto de vista pudiera oponerse seriamente que el juez robot no fuera a ser capaz de motivar de esa manera. Sin embargo, es cierto, que para motivaciones más complejas los sistemas de IA en la actualidad podrían prestar una extraordinaria labor asistencial, pero aun sería necesaria la supervisión de un juez humano que concretara la motivación.

En similar sentido se expresa Conde Fuentes (2022, p. 7) que afirma que

[...] si queremos que una decisión judicial automatizada sea respetuosa con la exigencia constitucional de motivación, creemos que el resultado que ofrezca no puede entrar en la motivación orientada al caso concreto; en todo caso, puede —y debe tan sólo— ofrecer una motivación tipo, general, que puede ser utilizada en diversos casos similares [...] esa motivación tipo, que puede ofrecer un algoritmo, tan sólo sería posible en casos menos complejos y similares entre sí, debiendo completarse —necesariamente— con la motivación referida al caso concreto y que sólo el juez puede realizar interpretando las normas aplicables.

Como señala Nieva Fenoll (2018, p. 103):

Imagínese lo que va a suponer conseguir optimizar una tarea que indudablemente también implica el lenguaje, hablado o escrito, pero que además tiene que tener en cuenta otros datos que debe engarzar con las declaraciones documentales o de los interrogatorios. La tarea es ciertamente titánica, pero cambiará la aproximación a las citadas pruebas y las sacará en gran medida del subjetivismo de cada juez.

2.3. La confiabilidad de la decisión: ¿hasta qué punto sería segura la decisión robótica? Imparcialidad e independencia

Como decíamos al principio de este trabajo, una vez comprobado que el juicio robótico sea posible, lo que había que determinar, a renglón seguido, es si es o no conveniente hacer uso de él.

Para eso habría que considerar toda una serie de obstáculos que se yerguen ante este tipo de decisiones y ver si podían o no ser removidos. Algunas de estas trabas las plantean las exigencias del derecho de defensa, las de la presunción de inocencia, las de la igualdad de armas o el mantenimiento de la imparcialidad y la independencia. Por razones de espacio voy a referirme en este trabajo sólo a las dos últimas.

La imparcialidad y la independencia, si lo pensamos bien, no son más que dos especialidades de una categoría genérica. Esa categoría genérica sería “decisiones tomadas con sujeción a la ley”.

La sujeción a la ley se predica tanto del juez independiente como del juez imparcial.

La diferencia está en que con la independencia se trata de enervar las influencias, extrañas a la ley, provenientes del sistema social que pueden producir decisiones judiciales desviadas. Mientras que con la imparcialidad se trata de enervar las influencias extrañas a la ley, pero provenientes del proceso y que pueden igualmente producir decisiones judiciales desviadas. Una decisión judicial desviada sería aquella tomada sin sometimiento pleno a la ley.

Desde este punto de vista, la categoría genérica “decisiones tomadas con sujeción a la ley” tendría dos especialidades: “las decisiones tomadas con imparcialidad” y “las decisiones tomadas con independencia”. O lo que es lo mismo, una decisión tomada con sujeción a la ley debería ser así ser ajena a influencias extrañas a la ley provengan éstas de donde provengan.

El juez independiente es aquel que depende, en el ejercicio de su función, exclusivamente de la ley y de ninguna otra influencia ajena a ella. El juez imparcial será aquel que no tiene ningún interés vinculado ni a las partes ni al objeto de un determinado proceso.

Pues bien, si examinamos la posible independencia e imparcialidad del juez robot, algunos autores no albergan dudas de que está fuera de la realidad (Gómez Colomer, *ibídem*, 2023, p. 217):

[...] hipotéticamente sería posible imaginar la absoluta independencia e imparcialidad del juez robot, en tanto en cuanto pueden ser programados con prevención frente a cualquier influencia externa o interna. Pero es irreal, porque precisamente la programación la hace el ser humano, que o bien es incapaz de lograr tan alto índice de objetividad, o bien no puede resistirse a sus propios sesgos.

En mi opinión, se insiste en sacralizar la figura del juez humano considerándolo, no invulnerable, pero sí más controlable en sus sesgos de lo que lo sería el juez robot. Estimo que es una visión distorsionada. Nadie sabe exactamente cuáles son los sesgos asentados en la mente de los jueces, pues si algo es seguro, es que los tienen, como seres humanos que son.

La diferencia con los sesgos robóticos es que éstos se exteriorizan claramente —en tanto en cuanto el algoritmo sea transparente y esté bien evaluado— y desde esa perspectiva es más fácil mitigarlos. Los sesgos humanos pueden, sin embargo, pasar desapercibidos a menos que se muestren a las claras.

Quienes parten de la imposibilidad de articular un juez robot imparcial e independiente y niegan por tanto su viabilidad, se aventuran afirmando que

quienes lo creen viable (Gómez Colomer, 2023, p. 220)

[...] no son juristas, o si lo son, no son procesalistas, y además son teóricos, no prácticos, aunque hablen siempre de la realidad social. Si fuesen procesalistas que observan atentamente esa realidad, lo verían mucho más claro.

Empero, la extracción profesional de algunos de los defensores de la figura del juez robot desmiente esas afirmaciones tan categóricas. Sin ir más lejos, los magistrados de la Audiencia Nacional (AN) Velasco Núñez y Delgado Martín —ambos miembros del Poder Judicial con dilatadas trayectorias y miembros a su vez del Grupo de expertos sobre IA del CGPJ—, sí atisban la existencia futura del juez robot como hemos tenido oportunidad de mostrar *supra* (9).

Para garantizar la imparcialidad e independencia de los sistemas de IA conformadores del juez robot, creemos que será necesario ejercer dos acciones.

En primer lugar, estimamos con Conde (2022) que será imprescindible la creación de un organismo público que controle los algoritmos de modo que se garantice la independencia judicial.

Ese organismo bien pudiera ser la Agencia Española de Supervisión de la Inteligencia Artificial (AESIA) con sede en La Coruña, puesto que su misión general es la de supervisar el cumplimiento de la normativa europea sobre IA en España, lo que indiscutiblemente incluye velar por la calidad de los criterios usados en los algoritmos de IA. Habrá que esperar a la ley que regule esa Agencia para comprobar si cabe entre sus funciones ésta que apuntamos.

En segundo lugar, como señala Nieva Fenoll (2018, p. 122), será también vital cuidar muy bien la contratación de los técnicos que elaboran los algoritmos

[...] de manera que se escoja a personas que estén ideológicamente orientadas a la defensa de los valores democráticos, en la medida en que ello sea averiguable sin entrar en situaciones discriminatorias, lo que no resulta nada fácil.

Este autor plantea realizar pruebas psicotécnicas sobre este personal, así como exámenes psicológicos periódicos. En nuestra opinión esta clase de pruebas también debería poder hacerse a los jueces humanos para comprobar a lo largo de los años de su desempeño si siguen estando en condiciones de juzgar de manera independiente.

2.4. Control de la actividad del juez robot

Puesto que, como hemos tenido oportunidad de analizar, existen todavía riesgos importantes que

rodean la actividad del juez robot, se pueden utilizar dos mecanismos para controlarla. Se trata de la conocida como reserva de humanidad y de la supervisión humana (Ponce Solé, 2022).

Con la primera se establecen determinadas materias que, por ahora, no podrán ser gestionadas por sistemas de IA por no poder garantizarse la seguridad al emplear esta tecnología o por ser abiertamente contrarias a los derechos de las personas. A través de la segunda se establece la fiscalización *a posteriori* de la actividad de los sistemas de IA.

La *Artificial Intelligence Act* utiliza estos dos mecanismos y establece cuatro niveles de riesgo en relación a ellos para los sistemas basados en IA: riesgo inaceptable y por lo tanto prohibido (reserva de humanidad), riesgo alto que será fiscalizado (supervisión humana), riesgo limitado y riesgo mínimo.

Los sistemas de IA con un nivel de riesgo inaceptable para la seguridad de las personas estarán estrictamente prohibidos en la Unión Europea. Esta será la lista de las prácticas de IA prohibidas en la UE (Gonzalo, 2023):

1. La identificación biométrica a distancia, como el reconocimiento facial, en tiempo real, en espacios de acceso público. Las organizaciones de derechos humanos consideran esta prohibición un paso positivo, ya que esta tecnología es demasiado invasiva y atenta contra derechos fundamentales.

Esta prohibición tiene algunas excepciones como cuando sea para buscar a víctimas de un delito, la prevención de una amenaza específica inminente para la vida o la seguridad física de las personas o de un atentado terrorista, o cuando se utilice para perseguir delitos con penas de prisión de más de tres años en el Estado miembro implicado.

2. La identificación biométrica a distancia de forma retroactiva, con la única excepción de las fuerzas de seguridad para la persecución de delitos graves y sólo previa autorización judicial.

3. La puntuación social (*social scoring*) cuando la utilicen entidades privadas, que es cuando un sistema de IA evalúa la fiabilidad de un individuo basándose en su comportamiento social o en sus preferencias, emociones, salud o inteligencia.

4. Sistemas policiales predictivos basados en perfiles, localización o comportamientos delictivos anteriores.

5. Sistemas de reconocimiento de emociones en las fuerzas de seguridad, la gestión de fronteras, el lugar de trabajo y las instituciones educativas.

6. Extracción indiscriminada de datos biométricos de redes sociales o grabaciones de CCTV para crear bases de datos de reconocimiento facial (violación de los derechos humanos y del derecho a la intimidad).

7. El uso de técnicas subliminales que trasciendan la conciencia de una persona para alterar su comportamiento de un modo que provoque perjuicios físicos o psicológicos a esa u otra persona.

8. La IA que aproveche alguna de las vulnerabilidades de un grupo específico de personas debido a su edad o discapacidad física o mental para alterar su comportamiento para provocar perjuicios físicos o psicológicos.

9. La utilización de sistemas de IA por parte de las autoridades públicas con el fin de evaluar o clasificar la fiabilidad de personas físicas durante un período determinado de tiempo atendiendo a su conducta social o a características de su personalidad conocidas o predichas, de forma que la clasificación social resultante provoque un trato perjudicial contra determinadas personas o colectivos en contextos sociales diferentes o que sea desproporcionado con respecto a su comportamiento social.

Como puede observarse algunas de estas prácticas prohibidas contemplan actividades que un juez robot podría utilizar para alcanzar su decisión como, por ejemplo, la identificación biométrica a distancia, la puntuación social o la clasificación de la fiabilidad de las personas. Parece, pues, razonable que todas estas actividades estén prohibidas.

En cuanto a los sistemas de alto riesgo entre ellos encontramos ciertos sistemas de IA destinados a la administración de justicia y los procesos democráticos, dado que pueden tener efectos potencialmente importantes para la democracia, el Estado de Derecho, las libertades individuales y el derecho a la tutela judicial efectiva y a un juez imparcial.

En particular, a fin de evitar el riesgo de posibles sesgos, errores y opacidades, procede considerar de alto riesgo aquellos sistemas de IA cuyo objetivo es ayudar a las autoridades judiciales a investigar e interpretar los hechos y el Derecho, y a aplicar la ley a unos hechos concretos.

En este caso, como decíamos, aplicamos el mecanismo de supervisión humana y habría que

atender a dos aspectos: quién supervisa y cómo se supervisa. Ambos aspectos son abordados por el Artificial Intelligence Act en su versión actual.

Respecto a quién supervisa, las personas físicas encargadas de garantizar la supervisión humana deberán tener un nivel suficiente de conocimientos de IA, así como el apoyo y la autoridad necesarios para ejercer dicha función, durante el período en que se utilice el sistema de IA y para permitir una investigación exhaustiva tras un incidente.

La norma especifica las capacidades que debe tener el sujeto supervisor. Así, según Ponce Solé (2022, p. 65)

[...] no basta con que haya un humano que vigile. Deber ser un humano cualificado y dotado de conocimientos y habilidades específicas para ello.

En cuanto a cómo se lleva a cabo la supervisión, se entiende que la supervisión permite:

1. Detectar y abordar lo antes posible los signos de anomalías, disfunciones y rendimiento inesperado.
2. Intervenir en el funcionamiento del sistema de IA de alto riesgo o interrumpirlo mediante un botón de "parada" o un procedimiento similar que permita que el sistema se detenga en un estado seguro, excepto si la interferencia humana aumenta los riesgos o tiene un impacto negativo en el rendimiento.

3. Reflexiones mirando al futuro

Tras haber examinado a lo largo de este artículo la capacidad del juez robot para cumplir la función jurisdiccional en su vertiente de declarar el derecho debemos concluir que puede hacerlo, pero con las limitaciones que se han indicado. Estas restricciones podrán irse difuminando conforme avance la tecnología quedando siempre las cuestiones más complejas sometidas al conocimiento de los jueces humanos.

En cuanto a la conveniencia de que el juez robot ejercite esa función jurisdiccional, desde el punto de vista de la imparcialidad y la independencia que son sólo dos de los obstáculos a considerar, la conclusión es también positiva con los matices indicados.

Desde esa posición en la que consideramos que el juez robot puede cumplir al menos una parte de la función jurisdiccional y en la que juzgamos que es conveniente que lo haga, cabe plantear algunas reflexiones mirando al futuro:

1. Es imprescindible que se regule el funcionamiento de los sistemas de IA en la Justicia.

Vistas las inmensas posibilidades de aplicación de la IA en el sector Justicia, es hora de que se regule su funcionamiento. En ese sentido será muy positiva la definitiva culminación de la *Artificial Intelligence Act* prevista para finales de 2023.

A partir de ese momento los Estados miembros de la UE tendrán que transponer la normativa a sus distintos ordenamientos nacionales y habrá que examinar cómo se lleva a cabo esa labor en nuestro país. El informe del grupo de expertos sobre IA del CGPJ ya apunta distintas posibilidades. Habrá que ver por cuáles se decanta nuestro legislador.

2. La aplicación de la IA en el sector Justicia no supondrá la desaparición del juez humano. Muy al contrario, los jueces humanos seguirán siendo pieza clave de nuestro sistema, pero se les liberará de la parte más tediosa de su trabajo reservándose sus capacidades para los asuntos más complicados y/o novedosos.
3. Habría que distinguir claramente entre los procesos reservados a la IA de aquellos otros objeto de juicio humano. La IA haría lo que pudiera resolver mejor (los procesos más sencillos, de menos cuantía, que no afecten a derechos fundamentales ni a personas vulnerables, etc.), mientras que los jueces humanos resolverían el resto de asuntos que serían de hecho los más sustanciales.
4. Debe garantizarse la revisión por jueces humanos de las decisiones tomadas por el juez robot y esto ha de ser concebido como un nuevo derecho de los justiciables. No sería solo ese derecho el que habría que añadir al elenco de los existentes, pero sí sería uno de los más capitales.
5. Debería implementarse un organismo público que controle los algoritmos de modo que se garantice la independencia judicial. Este organismo podría ser la AESIA y sabremos si lo es una vez se haga realidad la ley que la regule.
6. Deberían aprovecharse estas próximas reformas legislativas para dotar al proceso penal español de una nueva Ley de Enjuiciamiento Criminal largamente demandada.

Nuestra venerable Ley de Enjuiciamiento Criminal ya ha visto tres siglos y ha sido objeto de innumerables parcheos legislativos. Ha llegado la hora de sustituirla de una vez por todas. Resulta completamente anacrónico y contraproducente que una ley nacida a finales del siglo XIX incorpore reformas relativas a la aplicación de los sistemas de IA en el proceso penal.

Notas

- (1) La traducción de este documento y de todos los originales en inglés que aparecen en el texto, es nuestra.
- (2) Enmienda núm. 3 de la *Artificial Intelligence Act*.
- (3) Enmienda núm. 12 de la *Artificial Intelligence Act*.
- (4) En muchos casos cabría la revisión del caso en forma de recurso para hacer valer ante el tribunal superior la existencia de sesgo en la decisión recurrida.
- (5) Enmienda núm. 300 de la *Artificial Intelligence Act*.
- (6) Antes de que se hubiese alcanzado el texto común de la *Artificial Intelligence Act* del pasado 14 de junio de 2023.
- (7) Enmienda núm. 413 de la *Artificial Intelligence Act*.
- (8) STS 128/1995, de 9 de julio.
- (9) En el informe que dicho Grupo ha elaborado en 2022 hay un apartado dedicado a la toma de decisiones jurisdiccionales y, dentro de él, un subapartado de ayuda a la decisión automatizada y otro dedicado a los jueces robot. En este último se distinguen hasta tres clases del mismo: modelo Juez IA doctrinal; modelo Juez IA normativo y modelo Juez IA máximas de la experiencia.

Referencias

- Barona Vilar, S. (2021). *Algoritmización del Derecho y de la Justicia: De la Inteligencia Artificial a la Smart Justice*. Valencia: Tirant lo Blanch.
- Barrio Andrés, M. (2022). *Inteligencia artificial: origen, concepto, mito y realidad*. // *El Cronista del Estado social y democrático de derecho*. 00, septiembre-octubre.
- Comité Técnico Estatal de la Administración Judicial Electrónica (2022). *Manifiesto por un espacio público de datos en el ámbito de Justicia*. <https://www.mjusticia.gob.es/es/JusticiaEspana/ProyectosTransformacionJusticia/Documents/Manifiesto%20del%20Dato.pdf>
- Conde Fuentes, J. (2022) *Inteligencia artificial y robotización judicial: su impacto en nuestro sistema de justicia*. // *Derecho Digital e Innovación*. 13, julio-septiembre.
- Dalli, H. (2021). *Initial Appraisal of a European Commission Impact Assessment*. European Parliamentary Research Service. [https://www.europarl.europa.eu/RegData/etudes/BRIE/2021/694212/EPRS_BRI\(2021\)694212_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/BRIE/2021/694212/EPRS_BRI(2021)694212_EN.pdf)
- De la Oliva Santos, A. (2019). *Justicia predictiva, interpretación matemática de las normas, sentencias robóticas y la vieja historia del "Justizklavier"*. // *El Cronista del Estado social y democrático de derecho*. 100, mayo.
- Del Moral García, A. (2023). *Inteligencia artificial y Justicia penal, Comunicación en el Simposio Justicia y Derecho en Datos*. https://www.mjusticia.gob.es/es/AreaTematica/DocumentacionPublicaciones/InstListDownload/I_Simposio_Justicia_derecho_datos.pdf
- Delgado Martín, J. (2020). *Judicial-Tech, el proceso digital y la transformación tecnológica de la Justicia*. Madrid: La Ley.
- Delgado Martín, J. (2022). *La transformación digital de la justicia es un proceso que ya se ha iniciado pero en el que todavía queda mucho camino por recorrer*. // *Diario La Ley*, 17 de febrero. <https://dialnet.unirioja.es/servlet/articulo?codigo=9004665>
- Diario La Ley (2023). *El creador del ChatGPT solicita a los gobiernos que regulen la Inteligencia Artificial*. // *Diario La Ley*. 17 de mayo. https://diariolaley.laleynext.es/Content/DocumentoRelacionado.aspx?params=H4sIAAAAAAAEADWozWoDMQyEnya-Blp3N_076JKkh0IJSy9K15lbXAs17K38dvXJOQwoEGfZvRbKNWRrhkmSmQsr2NiQ4J-TRJxVbQ-d-yV1MChXmBMhVTGk0C3ejV906DQ5IJ-zwZeb5YtNOIJesWphW5rmzJn9EcSGIb-WYnlwMubsbsOGwx3WpDNMHh

- EuiVJ24J9LgxGg6K7efJNVQvkrdB8EI-7vzmu3NOALUgOnLsNHZgYzR-s-Lh_AJ3k50hrHMmip17yJTSgGmDx51zbPO_2zT2snLUC-BVHTsYwxXF6q71Lr0zZT-80MtbqslRG7khP5-ulK2b1e7D0-b-Dn6sfyPiFZk7h2vsLsLLH-vYAAAA=WKE
- Gonzalo, M. (2023). *Algoritma*. // *Newtral*. <https://www.newtral.es/ley-ia-union-europea-inteligencia-artificial-europa/20230520/>
- Gómez Colomer, J. L. (2021). *Unas reflexiones sobre el llamado juez-robot, al hilo del principio de la independencia judicial*. // Barona Vilar, Silvia (dir.); et. al. *Justicia algorítmica y neuroderecho: una mirada multidisciplinar*. Valencia: Tirant lo Blanch.
- Gómez Colomer, J. L. (2023). *El juez-robot: La independencia judicial en peligro*. Valencia: Tirant lo Blanch.
- Ignatieff, M. (2023). *El derecho a ser juzgados por humanos*. // *ABC*. 12 de febrero, 3.
- Marchena Gómez, M. (2022). *Discurso de ingreso en la Academia de doctores*. <https://confilegal.com/wp-content/uploads/2023/03/INTELIGENCIA-ARTIFICIAL-INGRESO-MARCHENA-REAL-ACADEMIA-DE-DOCTORES-.pdf>
- Martín Pastor, J. (2023). *Retos de la justicia digital*. // Jiménez Conde, Fernando; et al. *Logros y retos de la justicia civil en España*. Valencia: Tirant lo Blanch.
- Martínez Garay, L.; García Ortiz, Andrea (2022). *Paradojas de los algoritmos predictivos utilizados en el sistema de justicia penal*. // *El Cronista del Estado social y democrático de derecho*. 100, septiembre-octubre.
- Martínez Zorrilla, D. (2019). *El juez artificial ¿próxima parada?* // *Oikonomics*. 12, noviembre. <https://oikonomics.uoc.edu/divulgacio/oikonomics/es/numero12/dossier/dmartinez.html>
- Morales, O. (2023). *Los límites y desafíos de la inteligencia artificial*. // *Infobae.com*. <https://www.infobae.com/estados-unidos/2023/04/29/los-limites-y-desafios-de-la-inteligencia-artificial/>
- Nieva Fenoll, J. (2018). *Inteligencia artificial y proceso judicial*. Madrid: Marcial Pons.
- Peiró, K.; Baeza-Yates, R. (2022). *¿Pueden los algoritmos ser más justos que las personas?* // *CCCBLAB, Investigación e Innovación en Cultura*, 18 de enero. <https://lab.cccb.org/es/pueden-los-algoritmos-ser-mas-justos-que-las-personas/>
- Ponce Solé, J. (2022). *Reserva de humanidad y supervisión humana de la inteligencia artificial*. // *El Cronista del Estado social y democrático de derecho*. 100, septiembre-octubre.
- Rivas Velasco, M. J. (2023). *Uso ético de la inteligencia artificial en justicia*. // *Diario La Ley*, 13 de julio. <https://diariolaley.laleynext.es/dll/2023/07/13/uso-etico-de-inteligencia-artificial-en-justicia>
- Sanchis Crespo, C. (1999). *La prueba por soportes informáticos*. Valencia: Tirant lo Blanch.
- Sanchis Crespo, C. (2022). *Aspecto funcional de la Jurisdicción*. // Sanchis Crespo, C.; Pardo Iranzo, V.; Montón García, M. L.; Zaragoza Tejada, J. I. *Derecho Procesal I: Jurisdicción, Acción y Proceso*. Cizur Menor (Navarra): Aranzadi.
- Velasco Núñez, E. (2023). *La IA, con el control final humano, conllevará a una mejor justicia, más reflexionada y contextual* / A. J. Pérez Cruz, Entrevistador // *Diario La Ley*. 18 de mayo. https://diariolaley.laleynext.es/Content/Documento.aspx?params=H4sIAAAAAAAEADWozWoDMQyEnya-Blp3N_076JKkh0IJSy9K15lbXAs17K38dvXJOQwoEGfZvRbKNWRrhkmSmQsr2NiQ4J-TRJxVbQ-d-yV1MChXmBMhVTGk0C3ejV906DQ5IJ-zwZeb5YtNOIJesWphW5rmzJn9EcSGIb-WYnlwMubsbsOGwx3WpDNMHh

Q2vdd--bt02nFkrSAPhxM4VMYrrZfjXIOy-Eydhv-
nAlaebk0hp9Q4IUZ31Z7zLRDT2F6FGCMvh7Ztz9uPri8
Dj_DDhMXIQ_6HzIQACcUAQA AWKE

Velasco Núñez, E.; Cuatrecasas Monforte, C. S.; Buisan Gar-
cia, M. N.; Muñoz Aycuens, C.: Delgado Martín, J.; Ercilla
García, J.; Marquez Solís, S. (2022). Tecnología, IA y Ad-
ministración de Justicia: Catalogación de herramientas

tecnológicas para la Administración de Justicia (solu-
ciones tecnológicas emergentes en la Justicia). Informe
del grupo de expertos de trabajo del CGPJ sobre IA, no
publicado. [https://www.cej-mjusticia.es/sede/publicacion
es/ver/13637](https://www.cej-mjusticia.es/sede/publicacion
es/ver/13637)

Enviado: 2023-07-31. Segunda versión: 2023-10-03.
Aceptado: 2023-11-16.

A curatorial model for digital heritage collections as linked open data and named graphs

Un modelo curatorial para las colecciones de patrimonio digital como datos abiertos enlazados y grafos con nombre

Carlos H. MARCONDES

ECI- Information Science School, UFMG - Minas Gerais Federal University, PPGGOC - Graduate Program on Knowledge Management and Organization, Av. Pres. Antônio Carlos, 6627 - Pampulha, Belo Horizonte - MG, CEP 31270-901, Brazil, ch_marcondes@id.uff.br

Resumen

Un trabajo curatorial agrega valor al proporcionar una interpretación de dichos conjuntos de objetos, relacionando los objetos digitales patrimoniales entre sí y con otros recursos digitales, contextualizándolos, clasificándolos por tipología y brindando comentarios. Como tal, un trabajo curatorial es autorial. Se presenta una propuesta de modelo de curaduría de objetos patrimoniales digitales, con el objetivo de convertir el trabajo curatorial de colecciones de archivos, bibliotecas y museos en recursos digitales, permanentes, autorales y reutilizables. Actualmente, las colecciones de objetos patrimoniales se publican en la web como datos abiertos vinculados. El valor cultural y educativo del trabajo curatorial aumentaría si pudiera reutilizarse, como un catálogo de exhibición impreso. También se propone la implementación de dicho modelo utilizando LOD y tecnologías de grafos nombrados.

Palabras clave: Archivos. Bibliotecas. Museos. Colecciones. Trabajo curatorial. Modelos de patrimonio digital. Datos abiertos enlazados.

1. Introduction

Since the 2010s, several projects for publishing collections from heritage institutions as Linked Open Data (LOD) have begun to appear. Projects such as those of Europeana (1), those of the National Libraries of Spain (2), France (3) and Germany (4), those of the British Museum (5) are well known, and, more recently, of archival institutions, driven by the emergence of the new conceptual model Records in Contexts (RiC-CM) (2021), such as those of the National Archives of Sweden (6).

The profusion of digital resources of interest to Science, Culture and Art now available over the web enlarges the possibilities of using such resources for cultural and educational purposes. Applications such as digital libraries, virtual museums and exhibitions, e-learning and distance learning platforms are similar since users interact with these applications according to previously defined procedures (browsing through resource access options, searching for information, and accessing and interacting with resources). The

Abstract

Currently, collections of heritage objects are being published on the web as linked open data. A Curatorial work adds value by providing an interpretation of such sets of objects, relating heritage digital objects to each other and to other digital resources, contextualizing them, classifying them by typology, and providing commentary. As such a Curatorial work is authorial. A proposed model of curatorship for heritage digital objects is presented, with the goal of making the curatorial work on archives, libraries, and museum collections into digital, permanent, authorial, and reusable resources. The cultural and educational value of curatorial work would increase if it could be reused, like a printed exhibition catalogue. It is also proposed the implementation of such a model using LOD and named graph technologies.

Keywords: Archives. Libraries. Museums. Collections. Curatorial work. Digital heritage model. Linked open data.

proper functioning of such applications, be they digital libraries (Candela et al, 2007), virtual museums (Klentien, 2022) or e-learning and distance learning (Samsuzzaman et al., 2012), (Tunde et al., 2015), require conceptual models (Campos, 2004) to integrate different resources into coherent and user-friendly systems.

Digital collections from archives, libraries and museums are an important input for digital humanities research (Zeng, 2019). Nowadays objects in heritage collections have been published and represented on the Web using LOD technologies. Such collections have high cultural value, as they result from curatorial processes carried out by the institutions that have kept them permanently, such as archives, libraries, and museums. Furthermore, these collections published as LOD are open, not constituting information silos. Their digital objects can be used by anyone and interlinked with other resources.

Such collections can serve as inputs for research, science, education and culture. Their organiza-

tion, both in traditional finding aids such as catalogs, thesaurus and classification schemas, and in devices driven by information technologies such as ontologies, poses challenges to Knowledge Organization (Dalberg, 2014), Knowledge Representation (Giunchiglia, Dutta and Maltese, 2114) and to Conceptual Modeling (Guarino, 2009).

Many of these collections have thematic intersections or are related to other resources also available on the Web. For the initiatives of publishing these collections to reach their maximum cultural and educational impact, their digital objects should be integrated and interrelated with each other and with other entities represented as digital resources, such as authors, themes, events, periods histories and places. LOD technologies allow this integration and interoperability to be achieved (Marcondes, 2021) creating new and unusual navigational paths. However, such is not given “a priori”, but it requires curatorial work.

Relating heritage digital objects to each other and to other digital resources, classifying them by their typology and commenting on them is the result of a curatorial work *after* the publication of these collections as LOD. Such curatorial works can become permanent and reusable authorial resources in the form of curated virtual exhibitions, lectures, classes or collections.

This is ongoing research. The objective of this work is to develop a curatorial model for heritage collections (MIC) aimed at making the curatorial work into permanent, authorial and reusable resources. A use case implementation of such a model is also proposed using the LOD and named graphs technologies.

This article is organized as follows. After this introduction, where the research problem is established, the concept of curation of heritage collections, its phases and its role within the patrimonialization process discussed in section 2. Next, the current heritage models and the specifics of MIC are discussed in section 3. Section 4 presents, describes and discusses the proposed model. Finally, section 5 presents final remarks, conclusions and future works.

2. Digital curation of heritage collections

Today digital curation has a broad meaning encompassing activities from digital preservation to promoting access and reuse.

Heritage institutions such as archives, libraries, and museums play a vital role in curating collections of heritage objects and making these collections available to society for educational, cultural, and scientific purposes. Digital collections are

privileged material for the Digital Humanities (Zeng, 2019).

Being a heritage object is a cultural and value-added process that selects and separates objects from their original physical, functional, and cultural context and inserts them in a curatorial context: an archive, library, or museum. Such is the process of patrimonialization (7) where an object is selected, transferred from its original context, and brought to a heritage institution where its context is reconstructed. Within this new context, heritage objects are now witnesses and documents, representing or recording those aspects of the reality they came from (Van Mensh, 1992, 106), (Stransky, 1985, p. 98). Heritage objects are thus dual, they are primary objects (natural or man-made, or intangible) plus documents (Briet 2006)—descriptions of the primary object to add a semantic function and enrich its role as documents and testimony of natural, cultural, and social facts. As representations or symbols, i.e. documents, the properties assigned, added, or highlighted—their metadata (Qin and Zeng, 2020)—, during the curatorial process are dependent on the cultural, historical/social or scientific relevance of the specific object.

In other words, becoming a heritage object is a socially attributed characteristic (Searle, 1995), assigned to an object by an agent, a curator, on behalf of an archival institution, a library or a museum. This agent makes this decision based on the institution's collection policy and thus inscribes/lists the object in the institution's collection. The curator—be it an individual or an institution—thus performs an evaluation and decision-making process when carrying out the patrimonialization of a heritage object as stressed by Lima (2008, p. 36).

[...] valuer attribution, a judgment made by the cultural field that consigns it as an element with a differentiating character. And by distinguishing it in this way, it makes it 'special' and in a prominent position compared to other objects of the same nature, giving it a sense of exceptionality.

On the bases of such curatorial work developed by heritage institutions, others can be further developed. In the context of contemporary culture, a curator means, according to Gaskill (2011, p. 1), the professional in charge of collections in institutions who manage them, plans their exhibitions develop research to provide context.

LOD technologies allowed these collections to be published directly on the Web. Although heterogeneous, maybe represented using different metadata, conceptual models or vocabularies, LOD provides a unified data model. Once published, such collections are no longer limited to

the context of the original institutions that keep them; they can be combined with other entities represented on the Web such as authors, artists, themes, historical events, and places, in the ever-growing heritage interlinked “cloud”, forming new and richer resources.

During the 20th century, curatorship in culture shifted its focus from collections to exhibitions (O’Neil, 2007, p. 15). Currently such exhibitions are gaining a global dimension. Examples are “Da Vinci: Mechanics of Genius” at the Science Museum in London, curated by engineer Jim Bennett, “Da Vinci” at the Louvre Museum, Paris, which was principally curated by Vincent Delieuvin and Louis Frank, and “Human Bodies”, curated and catalogued by Prof. Gunther Von Hagens. With LOD technologies these exhibitions can now become new digital, authorial and permanent resources. Some of such are itinerant exhibitions that travelled by several countries, were visited by thousands of people, and were reassembled in different sites.

The instrument for disclosing and registering an exhibition is a catalog, a kind of special work. A catalog is a typical publication that permanently materializes the curatorship of a collection or exhibition. The content of an exhibition catalog are descriptions, analysis or comments about sets of heritage objects and why they are put together and related.

The Art History Research Guide (2023) of the University of Toronto Libraries stresses that

Exhibition catalogues are a very important type of literature for Art History. They provide documentation relating to all the items displayed in a show at a museum or art gallery and they contain new scholarly insight by way of thematic essays from curators and academics.

According to the same source, a catalog may have the following structure:

Now, exhibition catalogues often take the form of substantial books containing an introduction, essays, works shown, crisp colour images on glossy paper, a bibliography and sometimes an index.

An example of an online catalog, the British Museum catalogue, can be found at the following link: <https://www.britishmuseum.org/research/publications/online-research-catalogues>. The Guggenheim Museum, New York, makes available digitalized versions of all its catalogs in <https://archive.org/details/guggenheimmuseum>. A guide on writing exhibition catalogs can be found at https://owl.purdue.edu/owl/subject_specific_writing/writing_in_art_history/museum_catalog.html. An example of a printed catalog, the one of the “Da Vinci” exhibition at the Louvre Museum, Paris, can be found at <https://www.louvre.fr/en/what-s->

[on/exhibitions/leonardo-da-vinci#exhibition-catalogue](https://www.louvre.fr/en/what-s-on/exhibitions/leonardo-da-vinci#exhibition-catalogue).

In a catalog, a curator proposes a new look at that set of objects, contextualizes them, reveals new relationships—often unusual—between them, and proposes a narrative about those objects. Curating an exhibition is thus an authorial work. Therefore, the MIC aims somehow to mimic the structure and purposes of an exhibition catalog. It was developed based on the structure of the previously mentioned catalogs.

3. Current heritage models and the specifics of MIC

There are different conceptual models in the heritage domain as IFLA LRM, CIDOC CRM, BIBFRAME, EDM. Such models are declared as object-centred (Isaac, 2013, p. 12) or process/evento-centred (Isaac 2013, p. 15)

The aim of the CIDOC CRM is interoperability, i.e., “to allow exchange and sharing of information” (CRM Primer, 2012, p. 2) between heterogeneous data sources from different cultural heritage systems. It was designed to be implemented in RDF. It was conceived to model historical or cultural processes relating actors to heritage objects. In the CIDOC CRM no object is directly related to an actor, only through a process.

Another important remark is that the actor-event-object modelling used in the CIDOC CRM is quite complex and counterintuitive to be used by a curator. The CIDOC CRM also lacks a mechanism to provide a visiting sequence so useful within exhibitions.

The EDM—Europeana Data Model—, the conceptual model of the Europeana Digital Library, is a quite complex model. Europeana is a data aggregator. It has to keep linked different representations of the same heritage object provided by different heritage institutions (using the `ore:Aggregation` property), as, for example, the metadata about the Da Vinci’s Mona Lisa provided by both the Louvre Museum and the Joconde database (8), the catalog of objects of all France museums.

Within EDM, a heritage object may be related to persons, places, subjects, etc., that are part of its context, by two relationships: `edm:isRelatedTo` and `edm:hasMet` (Isaac 2013, p.13), depending if the modelling considers “object-centric” or “event-centric” approaches. The use of the EDM in such circumstances demands training and qualification by a curator.

Like the EDM, the MAP—Metadata Application Profile—, the conceptual model of the DPLA (DPLA Metadata Application Profile, 2017), the

Digital Public Library of America, also has to cope with different representations of the same heritage object provided by different partners.

The BIBFRAME–Bibliographic Framework–model is maintained by the US Library of Congress. It aims to provide the foundation for the future of bibliographic description, being an evolution of and resting on LC initiatives that date back to Project MARC in the 1970 decade. BIBFRAME was conceived to be serialized in RDF, being trivial the mapping of the bibliographic entity described in a MARC record–the Subject–of an RDF triple, the properties identified by MARC paragraphs/subfield codes–the triple’s Predicate–, and the content of such paragraphs/subfield codes–its Object.

Descriptive information in BIBFRAME is organized in three levels of abstraction–Work, Instance, and Item–, somehow simplifying the four levels of abstraction of models as the FRBR and CRM. The majority of properties provided by the BIBFRAME are descriptive properties of bibliographic entities or relationships between entities of different levels of abstraction.

The models FRBR, its successor–the LRM– and BIBFRAME, all rest on modelling relationships between levels of abstraction of the bibliographic entities aimed at aggregating all corresponding items of a collection in information retrieval tasks. FRBR and LRM have four levels of abstraction: work, expression, manifestation, and item; BIBFRAME has three: work, instance, and item. All such models have also relationships between bibliographic entities and entities as agents, concepts, time, and space.

One could say that the MIC is an object centred model, but in a special sense. It is best described as a *relationship-oriented model*. It aims *not* to provide properties to describe heritage objects but to use the descriptions *already* provided by heritage institutions, possibly using other models, vocabularies or ontologies, and to comment and relate such heritage objects representations one with another. To the MIC the important is the IRI through what a digital heritage object and its representation, no matter the model or vocabulary used to describe it, may be accessed.

Digital representations of heritage objects from collections in archives, libraries and museums are its focus. Curators assemble together sets of heritage objects, create narratives and a visiting path across such objects, and make statements about them using the proposed standard provided by the MIC and the CRR Vocabulary. Such statements are implemented as RDF triples. Agents (actors), themes, historical events and places provide context to heritage objects, being

also related to them using the CRR. Whenever possible the MIC uses other metadata vocabularies such as DC, SKOS, Schema.org, etc.

The MIC, through the CRR Vocabulary, relates DHOs that are both in the sense of the FRBR and LRM (2017) models Instances and Works. The Da Vinci’s Mona Lisa is both an Instance and a Work; the book “Mona Lisa: the history of the world’s most famous painting”, by Donald Sassoon is a Work with one (of several) Instance considered as a DHO within a case described with MIC. Such lack of ontological precision is a choice, aimed at making CRR intuitive and usable by curators without the need of any special training.

An item encompasses its expression, manifestation and work, but it, as a heritage object, has specific attributes of such levels of abstraction. MIC’s main focus are the relationships between such digital heritage objects, with each other and with other entities, for example, that Da Vinci’s Mona Lisa is 0012 Base_for the Dali’s self-portrait as Mona Lisa, not that the last is a manifestation of the former in the sense of FRBR, LRM, and BIBFRAME models.

Several research projects have focused on modeling narrative as a sequence of events describing a specific theme or domain. In the sequel are discussed proposals that are more closely related to the MIC.

Mulholland, Wolff and Collins (2012, p. 748) proposed the CURARE ontology. It aims not to model a specific heritage object but a curatorial narrative that

[...] spans across heritage objects and have a meaning that does not necessarily reside in the individual heritage objects themselves.

CURARE distinguishes between the conceptual structure of a story as a set of events comprising story entities, a plot as an interpretation of such story as causal relations, and a curatorial narrative, i.e., how such story is expressed. For example, in an exhibition catalogue story entities may be comprised of individual heritage objects narratives, a description about an object. The distinction made by CURARE between the structural aspects of a story and its expression enables the modelling of different exhibition catalogs to the same exhibition when, for example, it is a traveling exhibition presented in different spaces.

HiCo, the Ontology of Historical Context (Daquino and Tomasi, 2015), aims at formalizing as RDF triples, authoritative statements about heritage objects, specifically textual works, encoded according to TEI standards; such statements form another work, an authorial interpretation of the

original text that adds context to it. The original work and its interpretation by a curator are linked together through an interpretative act.

The use of LOD technologies to develop narratives and arguments is described in Renda and Daquino (2023).

Meghini, Bartalesi and Metilli (2021) claim that Digital Libraries (DLs), especially in the Cultural Heritage domain, are rich in narratives about the objects they hold. The authors propose the NOnt–Narrative Ontology–to formalize narratives within the context of a digital library. NOnt is the base of a Narrative Building and Visualising tool offered as a facility to users of a DL.

Among the different conceptual models, the one that has many similarities with the MIC is the Wikidata Data Model (9). Like MIC it holds statements, in the form of RDF triples, about the entity being described by a Wikidata page. Like the MIC uses the CRR Vocabulary to standardize statements about heritage objects, the statements in the Wikidata use the Wikidata:WikiProject Ontology (10) as Subjects, Predicates, and Objects of their RDF triples. Like the MIC, any statement in the Wikidata Data Model may be supported by a reference.

4. MIC–The Integrated Model of Curatorship

By curating heritage objects from different institutions published on the web as LOD, a digital curator performs tasks such as: (a) querying collections, (b) selecting DHOs, (c) gather and classifying sets of DHOs, (d) relating objects with each other and with other relevant heritage entities, (e) commenting and making assertions on them, (f) justifying such assertions by citing bibliographic or document sources.

The MIC aims to support such a curator's tasks. Although such tasks are processual, the MIC aims to model, not such processes, but their *recording* as data.

The MIC has as its focus the curatorial work–CW–as an authorial narrative. A CW is a subclass or specialization of the IFLA LRM-E2 Work. The core of MIC are the relationships provided by the CRR Vocabulary. They relate digital heritage objects–DHO–to each other and with other entities that make up the Heritage environment such as Agents, Concepts/Themes, Time, and Space.

When performing a curatorial task, a digital curator also gathers and classifies sets of objects according to criteria such as having the same creator (Agent), their connection to the same event or period (Time), with the same Concept/Theme or

with the same R 013 DHO_type, provided by the TOP Vocabulary (Marcondes, 2019). Such types may be assigned by himself using the CRR relationship Id: 025 Has_type or by the institution that publishes the DHO.

The relationships provided by the CRR vocabulary were obtained from cases suggested by cultural curators or obtained from the literature, as described in Marcondes (2020). Many of these relationships have similar ones provided by other vocabularies or conceptual models. However, in many such cases, there are no inverse relationships which are important to provide curators with flexible and intuitive modes of relation heritage objects with each other and with other entities. The CRR Vocabulary was conceived to be intuitive and concise, releasing the curator from training and study to use it.

Just as a scientific article has a special format (The Writing Center, 2023), the MIC is a model for a specific literary genre, the exhibition catalog. The MIC Model addresses the unique structural features of such a genre. Among them is the specification of the exhibition sequence of DHOs that make up the exhibition, making explicit the Culturally Relevant Relationships between them and between them and other entities, commenting on each of such features, and making citations to support them.

The MIC has the following

- Entities: Curatorial Works, CRR Annotations, Digital Heritage Objects identified in the class diagram by the yellow color);
- External entities: Agents, Concepts/Themes, Space and Time (E02 Agents, E02 Concepts/Themes, E04 Time, E05 Space, identified by the blue color).

The core of MIC classes is identified by gray color. Curators and Publishers classes (E021 Curators, E022 Publishers), subclasses of E02 Agents; digital Curatorial Works class (E08 Curatorial Works), with the same attributes of a publication. An important class is the E080 Sequence which functions as a summary specifying the sequence of E083 Cw_sections and E09 CRR_Annotations that make up the Curatorial Work. E080 Sequence is implemented as a skos:OrderedCollection construct.

The two types of parts of a Curatorial Work are the (1 to many) Curatorial Work sections class (E083 Cw_section), textual parts of a Curatorial Work, and the (1 to many) Annotations (E09 CRR_Annotations) class, each referring, i.e., using the Web Annotation Vocabulary (2017)

oa:hasTarget, to a single Culturally Relevant Relation class (R01 CRR_rel-Types) or any Heritage_Web_Resource. The last may be any external entity mentioned in a Curatorial Work or even a Heritage Object, especially detached in the Curatorial Work. A Curatorial Work is complemented with the Documents class, cited in a Section or Annotation (E082 Documents).

Relationships were reified in the representation as a class diagram and treated as classes (R01 CRR, identified in the class diagram by the pink colour), subdivided into two subclasses.

1. DHO Relationships/External Entities (R 011 DHO/Extern_ent_rel_types) as between a DHO and its author (R021 Created_by), between a DHO and an eventual contributor—the same as the Dublin Core metadata dc:contributor—(R 0051 Has_Contribution_of), between an DHO whose subject is an external entity (R 0061 Has_Subject), between an DHO and an

agent that has influenced its creation in some way (R 0071 Influenced), between an DHO and an agent associated with it, e.g. between the pen—the DHO—with which Brazilian Princess Isabel signed the Golden Law (11) and Princess Isabel—the agent—(R 010 Link_to_agent), between a DHO and its provenance as a place (R 0161 Provenance); and

2. DHO/DHO Relationships (R 012 DHO/DHO_rel_types), which can be seen in Figure 1. Some other relationships were used, as between DHO and external entities that are envisaged in other models or vocabularies, such as subject relationships between DHO and any of the external entities, authority/responsibility relationships between DHO and Agents, relationships such as dc:date (Dublin Core metadata vocabulary), between a DHO and the date assigned to it.

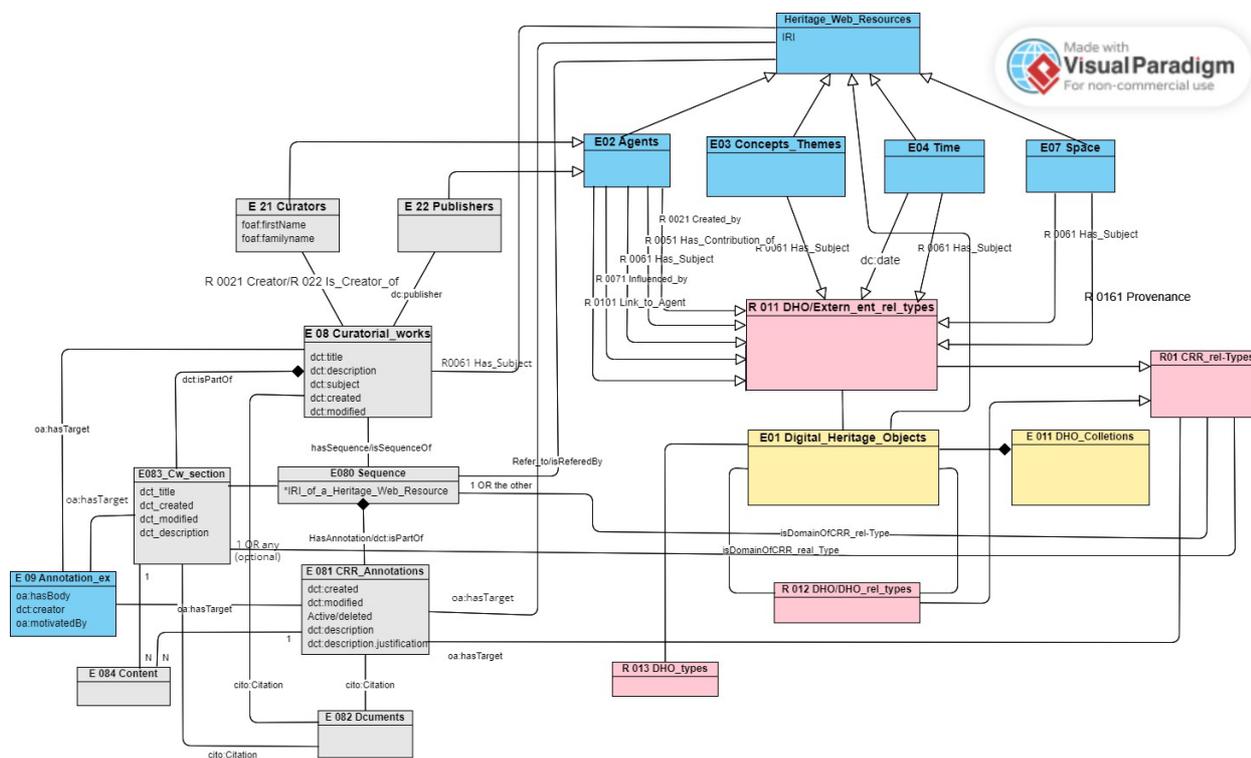


Figure 1. Class diagram of MIC

The attributes of the different classes of the MIC are only suggested and commented on when applicable.

In the proposed model, a CW has autonomy thus becoming a new document genre similar to an exhibition catalog. It becomes a subclass of a Work (FRBR, 1998; Riva et al., 2017; BIBFRAME, 2017) that has a Manifestation (FRBR, 1998; Riva et al., 2017) or Instance (BIBFRAME, 2017)

in a digital Item (FRBR, 1998; Riva et al., 2017), (BIBFRAME, 2017) with a specific IRI, where the author is identified, the digital Item is dated, and may contain other information that allow to identify their authorship and provenance.

An important issue in the work of a digital curator is to relate—by making links between them (Berners-Lee 2006)—DHO with each other and with other entities of interest to the culture represented

in the digital space. This is the purpose of CRR Vocabulary–Culturally Relevant Relationships (Marcondes 2020)–by providing a vocabulary to make these interrelationships. Using the CRR Vocabulary a curator could assert that the book “The Da Vinci Code”, by Dan Brown, inspire–crr:0081_inspired–the film of the same name

starring actor Tom Hanks, or asserts that the painting “Mona Lisa” by Da Vinci at the Louvre Museum served as the basis–crr:0012_base_for–the pastiche self-portrait of the Spanish painter Salvador Dalí portraying himself as the Mona Lisa. This case is illustrated in the Figure 2.



Figure 2. Da Vinci's Mona Lisa case as Linked Open Data

The CRR is implemented as predicates of a triple according to LOD technologies, having the URI of the DHO as the subject and the URI of another DHO or another entity as the object. The CRR is conceived to be intuitive e easy to be used by curators without the need of special training.

In the MIC, each CRR is treated as part of a CW, it is reified and becomes an annotation, receiving an IRI and also having its authorship and date recorded. Thus, a CRR_Annotation becomes an autonomous entity; it can be linked by citing another publication that justifies it and can receive other annotations as comments. It also has a status metadata, with two states, active/deleted, allowing the digital curator to review and modify a previously established CRR, so that readers can follow the changes made in the work. The MIC class diagram is presented in Figure 1. The CRR Vocabulary is described in Table I, on the right.

DHOs are digital representations of cultural heritage objects such as books, manuscripts, paintings, sculptures, artefacts in general, or manifestations of intangible culture or performing arts, such as a folk dance, a concert, or a ballet show, made permanent upon recorded and published on the Web.

Relationship	Inverse relationship
<i>Digital Heritage Object to Digital Heritage Object Relationships</i>	
Id: 0011 Based_on	Id: 0012 Base_for
Id: 0031 Design_or_Procedure_for	Id: 0032 Design_or_Procedure
Id: 0041 Documents	Id: 0042 Documented_by
Id: 0051 Has_Contribution_of	Id: 0052 Contributor
Id: 0071 Influenced_by	Id: 0072 Influenced
Id: 0081 Inspired	Id: 0082 Inspired_by
Id: 0091 Is_Illustrated_by	Id: 0092 Illustrated
Id: 0131 Mentioned_by_in	Id: 0132 Mentions
Id: 0141 Part_of	Id: 0142 Has_part
Id: 0151 Portrays	Id: 0152 Is_Portrayed_by
Id: 0171 Similar_item	
<i>Digital Heritage Object to External entities relationships</i>	
Id: 0021 Created_by	Id: 0022 Creator
Id: 0051 Has_Contribution_of	Id: 0052 Contributor
Id: 0061 Has_Subject	Id: 0062 Is_Subject
Id: 0071 Influenced	Id: 0072 Influenced_by
Id: 0101 Link_to_Agent	Id: 0102 Link_Agent_to_Object
Id: 0121 Link_to_Event_Process	Id: 0122 Link_Event_Process_to_Object
Id: 0161 Provenance	Id: 0162 Place_of_Provenance

Table I. The CRR Vocabulary (version 2.0)

These cultural manifestations are typified in the TOP–Classification of Types of Cultural Heritage Objects–(Marcondes 2019), also incorporated into the MIC.

5. Implementing MIC with LOD and named graphs technologies

LOD are enabling technologies to cultural heritage. They enable the development by culture curators of narratives (Meghini, Bartalesi, Metilli, 2021) historically situated in Space and Time, involving digital resources representing specific themes, heritage objects, events, agents. All such entities can be represented, interlinked, recoverable, and accessed. In addition, LOD and N-Graphs technologies can made up several interlinked digital resources into standalone, authorial and permanent resources.

The model presupposes that: (a) the CW are represented according to LOD technologies; (b) that the DHO already exist as such, are published on the Web by the institutions that curate the objects that the DHO represent, such as archives, librar-

ies, museums and others; and (c) it is also assumed that the digital curator already knows or has been notified of the existence of these DHO that he will use in his CW.

Named Graphs is an extension of RDF model in which a name, an identifier, as a link, is assigned to a set of triples (12) (Carrol, 2005). Thus, a graph can be identified and delimited, its navigation is not random, and can be planned according to an intended path. In addition, provenance and context information can be added.

The MIC is illustrated by a fictional case describing Da Vinci's Mona Lisa and its relations to resources describing Leonardo Da Vinci, the book and movie *The Da Vinci Code*, by Dan Brown, the Louvre Museum which holds the Mona Lisa, the Salvador Dali pastiche self-portrait as Mona Lisa, etc.

An implementation of MIC in a use case using LOD and named graph technologies is illustrated in the Appendix. Viewing Da Vinci's Mona Lisa graph through a browser can be seen in Figure 3 (using the GraphDB program).

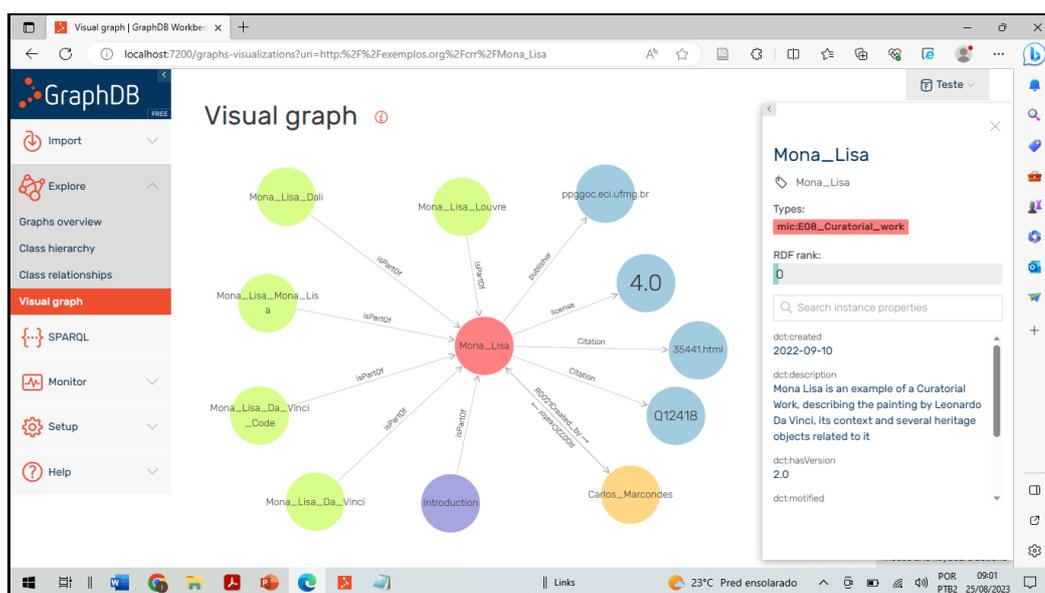


Figure 3. Viewing Da Vinci's Mona Lisa graph through a browser (using the GraphDB program)

5. Final remarks

To our best knowledge, there is no conceptual model in the heritage domain that addresses the modelling of a curatorial work. That is the aim of MIC.

This work outlined the MIC as a means of registering and making the work of digital curators autonomous and permanent by integrating digital objects from Heritage collections published on

the Web as LOD. To achieve these functionalities the MIC is implemented in RDF as a named graph. In this way, the curatorial work authorship of digital curators can constitute a new digital resource of educational and cultural value.

It is expected that the model will serve as a basis for the development of authorial tools for digital curation, or browsers through which users can browse DHOs, and curatorial works that integrate, annotate and comment on them.

The current version of the MIC can serve as a basis for the development of such applications.

As future developments, we intend to validate the MIC through new cases that represent significant curatorial works on Brazilian and global cultural heritage, load these cases into triple store and test the query possibilities using SPARQL. It is also intended to implement the MIC as a computational ontology.

Acknowledgments

This work was carried out with the support of the Brazilian agencies CAPES - Financing Code 001, and CNPq, grant number 303379/2022-7.

Notes

- (1) See <https://pro.europeana.eu/page/linked-open-data>.
- (2) See datos.bn.es.

- (3) See <https://data.bnf.fr/>.
- (4) See https://www.dnb.de/EN/Professionell/Metadatendienste/Datenbezug/LDS/lids_node.html.
- (5) See <https://old.datahub.io/dataset/british-museum-collection>.
- (6) See https://groups.google.com/g/records_in_contexts_users/c/KZQXeodod9k/m/SQ5tblsPAAAJ.
- (7) See <https://www.igi-global.com/dictionary/educommunication-web-20-for-heritage/82076>.
- (8) See <https://www.culture.gouv.fr/Espace-documentation/Bases-de-donnees/Fiches-bases-de-donnees/Joconde-catalogue-collectif-des-collections-des-musees-de-France>.
- (9) See <https://www.mediawiki.org/wiki/Wikibase/DataModel/Primer>.
- (10) See https://www.wikidata.org/wiki/Wikidata:WikiProject_Ontology.
- (11) See https://en.wikipedia.org/wiki/Lei_Áurea.
- (12) See <https://www.w3.org/2009/07/NamedGraph.html>.

Appendix: Implementation of the MIC as LOD and Named Graph

The coding of MIC as LOD and Named Graph, as follows, was validated by the IDLab Turtle validator, <http://ttl.summerofcode.be/>.

```
## A Curatorial Work relating Da Vinci's Mona Lisa with other heritage digital resources
@prefix dct: <http://purl.org/dc/terms/> .
@prefix crr: <http://exemplos.org/crr/> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix mic: <http://exemplos.org/mic/> .
@prefix top: <http://exemplos.org/top/> .
@prefix schema: <http://schema.org/> .
@prefix foaf: <http://xmlns.com/foaf/0.1/> .
@prefix skos: <http://www.w3.org/2004/02/skos/core#> .
@prefix cito: <http://purl.org/spar/cito/> .
@prefix wo: <https://www.wikidata.org/wiki/> .
@Prefix oa: <http://www.w3.org/ns/oa#> .

# Curatorial Work relating the Da Vinci Mona Lisa with other digital resources
<http://exemplos.org/crr/Mona_Lisa> a mic:E08_Curatorial_work;
  dct:title "Mona Lisa, an example of a Curatorial Work ";
  dct:description "Mona Lisa is an example of a Curatorial Work, describing the painting by Leonardo Da Vinci, its context and several heritage objects related to it";
  dct:hasVersion "2.0";
  dct:created "2022-09-10";
  dct:modified "2022-09-13";
  crr:R0021Created_by <http://exemplos.org/mic/Carlos_Marcondes>;
  dct:license <https://creativecommons.org/licenses/by-nc-nd/4.0/>;
  dct:subject "Italian Renaissance";
  dct:subject "Paintining";
  dct:publisher <https://ppggoc.eci.ufmg.br/>;

## Citations in the Curatorial Work
# Mona Lisa in the Wikidata
cito:Citation <https://www.wikidata.org/entity/Q12418>;
# Louvre Museum Catalog
cito:Citation <https://www.boutiquesdemusees.fr/en/exhibition-catalogues/monna-lisa-exhibition-catalogue/35441.html>.

## Annotations
# Mona Lisa related to Leonardo Da Vinci
mic:Mona_Lisa_Da_Vinci rdf:type skos:Concept.
# Mona Lisa related to The Da Vinci Code, by Dan Brown
mic:Mona_Lisa_Da_Vinci_Code rdf:type skos:Concept.
# Mona Lisa related to the Salvador Dali pastiche self-portrait as Mona Lisa
mic:Mona_Lisa_Dali rdf:type skos:Concept.
# Mona Lisa related to Louvre Museum
mic:Mona_Lisa_Louvre rdf:type skos:Concept.

## Sections
# Introductory Section
mic:Mona_Lisa_Introduction rdf:type skos:Concept.

## Sequence within the Curatorial Work
<Mona_Lisa_sequenceOfResources> rdf:type skos:OrderedCollection ;
  skos:prefLabel "Sequence of heritage web resources referring to Da Vinci's Mona Lisa"@en;
  skos:memberList <http://exemplos.org/mic/Mona_Lisa_Da_Vinci>.

mic:Mona_Lisa_Introduction rdf:first <http://exemplos.org/mic/Mona_Lisa/Introduction>;
  rdf:rest mic:Mona_Lisa_Da_Vinci.
mic:Mona_Lisa_Da_Vinci rdf:first <http://exemplos.org/mic/Mona_Lisa_Da_Vinci>;
  rdf:rest mic:Mona_Lisa_Da_Vinci_Code.
mic:Mona_Lisa_Da_Vinci_Code rdf:first <http://exemplos.org/mic/Mona_Lisa_Da_Vinci_Code>;
  rdf:rest mic:Mona_Lisa_Dali.
mic:Mona_Lisa_Dali rdf:first <http://exemplos.org/mic/Mona_Lisa_Dali>;
  rdf:rest mic:Mona_Lisa_Louvre.
mic:Mona_Lisa_Louvre rdf:first <http://exemplos.org/mic/Mona_Lisa_Louvre>;
```

```

rdf:rest rdf:nil .

# Mona Lisa
<https://collections.louvre.fr/en/ark:/53355/cl010062370> rdf:type schema:Painting ;
dct:type top:Painting;
dct:title "Mona Lisa";
dct:date "1503";
crr:0065InCollection <https://www.louvre.fr>;
crr:0131Mentioned_by_in <https://bnb.data.bl.uk/doc/resource/006946635>;
crr:0012Base_for <http://www.studiolo.org/Mona/MONASV12.htm>;
wo:P18 <https://commons.wikimedia.org/wiki/File:Mona_Lisa_by_Leonardo_da_Vinci,_from_C2RMF_retouched.jpg> .

# Leonardo Da Vinci
<https://www.wikidata.org/entity/Q762> rdf:type schema:artist;
dct:title "Leonardo da Vinci";
wo:image:P18 <https://www.wikidata.org/wiki/Q762#/media/File:Francesco_Melzi_-_Portrait_of_Leonardo.png>.

# The Da Vinci Code book
<https://bl.natbib-lod.org/the-da-vinci-code-p1121681401307175/> rdf:type schema:Ebook ;
dct:type top:textual_documents;
dct:title "The Da Vinci Code, by Dan Brown";
dct:creator "Dan Brown".

# The Louvre Museum
<https://www.louvre.fr> rdf:type schema:Museum;
dct:title "The Louvre Museum";
wo:P18 <https://commons.wikimedia.org/wiki/File:Louvre_2007_02_24_c.jpg>.

# the Salvador Dali pastiche self-portrait as Mona Lisa
<http://www.studiolo.org/Mona/MONASV12.htm> rdf:type schema:Painting;
dct:type top:Painting;
dct:title "Dali as Mona Lisa";
dct:creator "Salvador Dali";
wo:P18 <https://www.dalipaintings.com/self-portrait-mona-lisa.jsp#prettyPhoto[imagem1]/0/>.

# Carlos Marcondes
<http://exemplos.org/mic/Carlos_Marcondes> rdf:type mic:E21_Curator;
foaf:family_name "Marcondes";
foaf:first_name "Carlos";
foaf:homepage <http://profmarcondes.org.br> ;
crr:R0022Creator <http://exemplos.org/crr/Mona_Lisa>;
foaf:mbox <ch_marcondes@id.uff.br>.

## Annotations
<http://exemplos.org/mic/Mona_Lisa_Da_Vinci> rdf:type mic:E081_Annotation;
mic:domain_resource <https://collections.louvre.fr/en/ark:/53355/cl010062370>;
crr:0021Creator <https://www.wikidata.org/wiki/Q762>;
dct:isPartOf <http://exemplos.org/crr/Mona_Lisa>;
mic:A08_Content <http://exemplos.org/Text1>;
# The Wikidata page of Mona Lisa
cito:Citation <https://www.wikidata.org/wiki/Q12418>;
dct:date "2022-09-10".

<http://exemplos.org/mic/Mona_Lisa_Da_Vinci_Code> rdf:type mic:E081_Annotation;
mic:domain_resource <https://collections.louvre.fr/en/ark:/53355/cl010062370>;
crr:0131Mentioned_by_in <https://bnb.data.bl.uk/doc/resource/006946635>;
dct:isPartOf <http://exemplos.org/crr/Mona_Lisa>;
mic:A08_Content <http://exemplos.org/Audio1>;
dct:date "2022-09-10".

<http://exemplos.org/mic/Mona_Lisa_Dali> rdf:type mic:E081_Annotation;
mic:domain_resource <https://collections.louvre.fr/en/ark:/53355/cl010062370>;
crr:0012Base_for <http://www.studiolo.org/Mona/MONASV12.htm>;
dct:isPartOf <http://exemplos.org/crr/Mona_Lisa>;
mic:A08_Content <http://exemplos.org/Video1>;
dct:date "2022-09-10".

<http://exemplos.org/mic/Mona_Lisa_Louvre> rdf:type mic:E081_Annotation;
mic:domain_resource <https://collections.louvre.fr/en/ark:/53355/cl010062370>;
crr:0065InCollection <https://www.louvre.fr>;
dct:isPartOf <http://exemplos.org/crr/Mona_Lisa>;
# The Louvre Museum Catalog
cito:Citation <https://www.boutiquesdemusees.fr/en/exhibition-catalogues/monna-lisa-exhibition-catalogue/35441.html>;
dct:date "2022-09-10".

## Sections
# Introductory Section
<http://exemplos.org/mic/Mona_Lisa/Introduction> rdf:type mic:E083_CW_section;
dct:title "Da Vinci's Mona Lisa as seen by other artists";
dct:creator "Carlos H. Marcondes";
dct:isPartOf <http://exemplos.org/crr/Mona_Lisa>;
dct:description "The Da Vinci's Mona Lisa is the most famous painting of Occidental Art...";
dct:description <http://exemplos.org/any_text>;
dct:date "2022-09-10".

```

References

- Art History. Research Guide (2023). University of Toronto Libraries. (2023). <https://guides.library.utoronto.ca/art> (2023-06-22)
- Berners-Lee, Tim. (2006). Linked Data. Design Issues. W3C. // International Journal on Semantic Web and Information Systems. 4 (2006).
- BIBFRAME – Bibliographic Framework: Overview of the BIBFRAME 2.0 Model. (2017). Library of Congress.
- (2017). <https://www.loc.gov/bibframe/docs/bibframe2-model.html>.
- Briet, S. (2006). What is documentation? English translation of the classic French text. Scarecrow Press.
- Campos, Maria Luiza de Almeida. (2004). Modelling of knowledge domains: an investigation of fundamental principles. // Ciência da Informação 33 (2004) 22-32.
- Candela, L; Castelli, D; Ferro, N; Ioannidis, Y; Koutrika, G; Meghini, C; [...] Schuldt, H. (2007). The DELOS digital library reference model. Foundations for digital libraries.

- Carroll, Jeremy J.; et al. (2005). Named graphs. // *Journal of Web Semantics*. 3:4 (2005) 247-267.
- Dahlberg, I. (2014). What is knowledge organization?. // *Knowledge Organization*. 41:1, 85-91.
- DPLA Metadata Application Profile, 5. (2017). DPLA. <http://dp.la/info/map>
- FRBR - Functional requirements for bibliographic records: final report (1998). IFLA Study Group on the Functional Requirements for Bibliographic Records. München: K.G. Saur, 1998.
- Gaskill, Karen. (2011). Curatorial cultures: Considering dynamic curatorial practice. <http://shura.shu.ac.uk/4441/1/kgaskill.pdf>.
- Giunchiglia, F; Dutta, B; Maltese, V. (2014). From knowledge organization to knowledge representation. // *Knowledge Organization* 4:1, 44-56.
- Guarino, N. (2009). The ontological level: Revisiting 30 years of knowledge representation. // *Conceptual modeling: Foundations and applications*. Springer, Berlin, Heidelberg. 52-67.
- Isaac, A. (2013). *Europeana Data Model Primer*. Europeana. (2013). https://pro.europeana.eu/files/Europeana_Profesional/Share_your_data/Technical_requirements/EDM_Documentation/EDM_Primer_130714.pdf.
- Lima, D. F. C. (2008). Herança cultural (re)interpretada ou a memória social e a instituição museu: releitura e reflexões. // *Revista Museologia e Patrimônio*. Rio de Janeiro. 1:1, 3-43. <http://revistamuseologiaepatrimonio.mast.br/index.php/ppgpmus/article/view/%204/2> (2022-05-23).
- Marcondes, C. H. (2020). Towards a Vocabulary to Implement Culturally Relevant Relationships between Digital Collections in Heritage Institutions. // *Knowledge Organization*. 47:2 122-137 (2020).
- Marcondes, C. H. (2019). Una clasificación de tipos de objetos de patrimonio para la integración de acervos digitales de archivos, bibliotecas y museos. // *Scire*. 25:2 45-52 (2019).
- Marcondes, C. H. (2021). Integrated classification schemas to interlink cultural heritage collections over the web using LOD technologies. // *International Journal of Metadata, Semantics and Ontologies*. 15:3 (2021).
- Meghini, C; Bartalesi, V; Metilli, D. (2021). Representing narratives in digital libraries: The narrative ontology. // *Semantic Web*. 12:2, 241-264.
- Mulholland, P; Wolff, A; Collins, T. (2012). Curate and storyspace: an ontology and web-based environment for describing curatorial narratives. // *The Semantic Web: Research and Applications: 9th Extended Semantic Web Conference, ESWC 2012, Heraklion, Crete, Greece, May 27-31, 2012*. /Springer Berlin Heidelberg. 748-762
- Renda, G; Daquino, M. (2023). Storytelling with Linked Open Data. // *La Memoria Digitale, AIUCD Associazione per l'Informatica Umanistica e la Cultura Digitale*. 164-167.
- Ric-CM - Records in Contexts Conceptual Model. (2021) . ICA-EGAD. <https://www.ica.org/en/records-in-contexts-conceptual-model> 2023-09-26.
- Riva, Pat; Le Boeuf; Patrick; Žumer, Maja. (2017) IFLA Library Reference Model: A Conceptual Model for Bibliographic Information. IFLA. https://www.ifla.org/files/assets/cataloguing/frbr-lrm/ifla-lrm-august-2017_rev201712.pdf.
- Samsuzzaman, M; Islam, M; Rashid, S; Ahmed, F; Khan, M. (2012). Proposed model of e-learning management system using semantic web. // *Journal of Applied Sciences Research*. 8:5, 2484-2492.
- Klentien, U. (2022). Development of virtual museum model for youth. // *International Journal of Information and Education Technology*. 12:4, 313-317.

Enviado: 2023-03-29. Segunda versión: 2023-09-29.

Aceptado: 2023-11-22.

Índice de autores

Author index

Almeida, Carlos Cândido, 25
Ardines, Saul, 51
Baptista, Ana Alice, 39
De León, Amarilis, 51
Delgado Vázquez, Ángel María, 39
Juárez, Virginia, 51
López Carreño, Rosana, 39
Marcondes, Carlos H., 85

Martines, Alexandre Robson, 25
Martínez Méndez, Francisco Javier, 39
Paz, María Luisa, 51
Perdomo, Isaac, 51
Sanchis Crespo, Carolina, 65
Suárez Sánchez, Adriana, 13
Zazo, Ángel, 51

Índice de materias en español

Subject index in Spanish

Análisis de redes sociales, 51
Análisis reticular de coincidencias, 51
ARC, 51
Archivos, 85
ARS, 51
Bibliometría, 13
Bibliotecas, 85
Ciencia abierta, 39
Colecciones, 85
Conjuntos de datos de investigación, 39
Datos abiertos enlazados, 85
Desechos tecnológicos, 51
España, 39
Fundamentos lingüísticos, 25

Indización, 25
Lingüística, 25
Modales, 25
Modelos de patrimonio digital, 85
Museos, 85
Ontologías terminológicas, 13
Organización del conocimiento, 13
Panamá, 51
Producción científica, 13
RAEE, 51
Repositorios de datos de investigación, 39
Trabajo curatorial, 85
Universidades públicas, 39

Índice de materias en inglés

Subject index in English

Archives, 85
Bibliometrics, 13
Collections, 85
Curatorial work, 85
Digital heritage model, 85
E-Waste, 51
Indexing, 25
Knowledge organization, 13
Libraries, 85
Linguistic fundamentals, 25
Linguistics, 25
Linked open data, 85
Modals, 25
Museums, 85

NAC, 51
Network analysis of coincidences, 51
Open science, 39
Panamá, 51
Public universities, 39
Research data repositories, 39
Research datasets, 39
Scientific research, 13
SNA, 51
Social network analysis, 51
Spain, 39
Terminological ontologies, 13
WEEE, 51