





# ***SCIRE***

Representación y organización del conocimiento



# SCIRE

Representación y organización del conocimiento

Vol. 27, n.º 1, enero-junio 2021

ISSN 1135-3716

ISSN (e) 2340-7042

Scire:

knowledge representation and organization

Vol. 27, n. 1, January-June 2021

**Número especial:**  
Organización del conocimiento,  
patrimonio cultural y turismo  
= Knowledge organization,  
cultural heritage and tourism

Ibersid:

Red de Investigación  
en Sistemas de Información  
y Documentación

© 2021 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

---

Las instrucciones actualizadas para presentación de trabajos se encuentran en <https://www.iber-sid.eu/ojs/index.php/scire/about/submissions>; y del procedimiento de evaluación editorial en <https://www.iber-sid.eu/ojs/index.php/scire/about>.

## Agradecimientos

---

Este número de la revista *Scire*, dedicado a la organización del conocimiento para la interoperabilidad del turismo y patrimonio cultural se ha financiado con fondos del proyecto CSO2021-65448-R (MINECO/FEDER) y la ayuda del Vicerrectorado de Investigación y de 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](mailto: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 is available at <https://www.iber-sid.eu/ojs/index.php/scire/about/submissions>; and of the assessment process at <https://www.iber-sid.eu/ojs/index.php/scire/about>.

## Acknowledgments

---

This special issue of *Scire*, devoted to knowledge organization for the interoperability of cultural tourism and cultural heritage is founded by the project CSO2021-65448-R (MINECO/FEDER) and the help of the Research Vice Rectorate and the Faculty of 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](mailto: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 Montevideo, Uruguay	João Batista Ernesto Moraes, Universidade Estadual Paulista, Brasil	Francisco Javier Martínez Mendez, Universidad de Murcia, España
José Augusto Chaves Guimarães, Universidade Estadual Paulista, Brasil	Francisco Javier García Marco, Universidad de Zaragoza, España	Álvaro Quijano Solís, Colegio de México, México

### Consejo científico / Scientific council

Dr. Isidro Aguillo Caño, IPP- CSIC, España	Vicente Pablo Guerrero Bote, Univ. de Extremadura, España	José Luis Otaí, Universidad Jaume I de Castellón, España
Tomás Baiget, EPI S. A., España	Víctor Herrero Solana, Univ. de Granada, España	Manuel José Pedraza Gracia, Universidad de Zaragoza, España
José Luis Bonal Zazo, Univ. de Extremadura, España	José María Izquierdo Arroyo, Universidad de Murcia, España	María Pinto Molina, Universidad de Granada, España
Mercedes Caridad Sebastián, Universidad Carlos III de Madrid, España	María Pilar Lasala Calleja, Universidad de Zaragoza, España	Gloria Ponjuán Dante, Universidad de La Habana, Cuba
Alberto Carreras Gargallo, Universidad de Zaragoza, España	Alfonso López Yepes, Universidad Complutense de Madrid, España	José Vicente Rodríguez Muñoz, Universidad de Murcia, España
Constança Espelt Busquets, Universidad de Barcelona, España	José López Yepes, Universidad Complutense de Madrid, España	Adelaida Román Román, CINDOC (Madrid), España
Juan Carlos Fernández Molina, Univ. de Granada, España	Pedro Marijuán Fernández, Universidad de Zaragoza, España	Juan Ros García, Universidad de Murcia, España
María Eulalia Fuentes Pujol, Universidad Autónoma de Barcelona, España	Bonifacio Martín del Brío, Universidad de Zaragoza, España	Francisco José Ruiz de Mendoza Ibáñez, Universidad de La Rioja, España
Fernando Galindo Ayuda, Universidad de Zaragoza, España	José Antonio Moreira González, Universidad Carlos III de Madrid, España	Félix Sagredo Fernández, Universidad Complutense de Madrid, España
Blanca Gil Urdiciáin, Universidad Complutense de Madrid, España	Purificación Moscoso Castro, Universidad de Alcalá, España	Elías Sanz Casado, Universidad Carlos III de Madrid, España
Isidoro Gil Leiva, Universidad de Murcia, España	Félix Moya Anegón, Universidad de Granada, España	Carlos Serrano Cinca, Universidad de Zaragoza, España
Alan Gilchrist, Cura Consortium, Reino Unido	María del Carmen Negrete Gutiérrez, Universidad Autónoma de México	<b>In memoriam:</b> Emilia Currás Puente Emilio Setién Quesada

### Revisores externos del número / External reviewers in this issue

Agradecemos la colaboración generosa y desinteresada de, además de los miembros del consejo científico, María Luisa Alvite Díaz, Carlos G. Figuerola, Luis Martínez Uribe, João Batista Ernesto de Moraes, Javier Noguera Iso, Paloma María Santos, Javier Tarango Ortiz, y Ángel Zazo Rodríguez.

#### 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

*El turismo cultural y la digitalización  
del patrimonio en un entorno Linked Open  
Data: el caso de Europeana*

Xavier AGENJO BULLÓN  
Francisca HERNÁNDEZ CARRASCAL.....13

*Enriquecimiento automático de portales  
culturales mediante modelos de organización  
del conocimiento*

Dayany DÍAZ-CORONA  
Javier LACASTA  
Javier NOGUERAS-ISO .....23

*Contar historias con los datos: Aragón Open  
Data Focus, una experiencia innovadora  
de reutilización de los datos del sector público*

Lorea AKERRETA ESCRIBANO  
Julián MOYANO COLLADO .....31

*La indización social en los catálogos digitales  
de GoodReads y Skoob*

Jessica Beatriz TOLARE  
Mariângela Spotti Lopes FUJITA  
Bruna Daniele de Oliveira SILVA .....45

*La organización del conocimiento al servicio  
de la interconexión entre las instituciones  
de la memoria y el turismo cultural*

Francisco Javier GARCÍA MARCO .....55

*Redes y comunidades de descriptores  
en artículos de Biblioteconomía y Ciencia  
de la Información (1971-2020): análisis  
de su evolución temporal mediante  
Técnicas de Análisis de Redes*

Carlos G. FIGUEROLA  
Modesto ESCOBAR MERCADO  
Ángel ZAZO RODRÍGUEZ  
José Luis ALONSO BERROCAL.....71

## Índices

Índice de autores.....85

Índice de materias en español.....85

Índice de materias en inglés.....85



---

# Tabla de contenidos en inglés

*Table of contents in English*

---

*Table of contents in Spanish* .....9

*Table of contents in English* .....11

## Articles

*Cultural tourism and heritage digitization  
in a Linked Open Data environment:  
the Europeana case*

Xavier AGENJO BULLÓN  
Francisca HERNÁNDEZ CARRASCAL.....13

*Automatic enrichment of cultural portals  
with knowledge organization systems*

Dayany DÍAZ-CORONA  
Javier LACASTA  
Javier NOGUERAS-ISO .....23

*Data storytelling: Aragón Open Data Focus  
an innovative experience of reusing public  
sector data*

Lorea AKERRETA ESCRIBANO  
Julián MOYANO COLLADO .....31

*Social indexing in GoodReads  
and Skoob digital catalogs*

Jessica Beatriz TOLARE  
Mariângela Spotti Lopes FUJITA  
Bruna Daniele de Oliveira SILVA .....45

*Knowledge organization for the interoperability  
among memory institutions and cultural  
tourism*

Francisco Javier GARCÍA MARCO .....55

*Descriptor networks and communities  
in Library and Information Science articles  
(1971-2020): analysis of their temporal  
evolution through Network Analysis  
Techniques*

Carlos G. FIGUEROLA  
Modesto ESCOBAR MERCADO  
Ángel ZAZO RODRÍGUEZ  
José Luis ALONSO BERROCAL.....71

## Indexes

*Author index* .....85

*Subject index in Spanish* .....85

*Subject index in English* .....85



---

# El turismo cultural y la digitalización del patrimonio en un entorno Linked Open Data: el caso de Europeana

*Cultural tourism and heritage digitization in a Linked Open Data environment: the Europeana case*

---

Xavier AGENJO BULLÓN (1), Francisca HERNÁNDEZ CARRASCAL (2)

(1) Fundación Ignacio Larramendi, c/Alenza 4, 28003 Madrid (España), xavier.agenjo@larramendi.es.

(2) Biblioteca Nacional de España, Paseo de Recoletos, 20-22, 28001 Madrid (España), mariaf.hernandez@bne.es.

## Resumen

El texto repasa los proyectos desarrollados por Europeana para mostrar el papel que la digitalización y los datos abiertos vinculados de patrimonio cultural pueden jugar en el fortalecimiento del sector del turismo cultural. Se aportan una serie de datos estadísticos básicos acerca del turismo y del turismo cultural para enmarcarlo como sector económico, y se revisan las estrategias fundamentales de las instituciones europeas sobre turismo cultural, así como el papel preponderante que le asignan a Europeana. Estas estrategias se han trastocado y acelerado en respuesta a la debacle económica provocada por la pandemia de COVID-19. Se mencionan los principales planes estratégicos del Gobierno de España para el desarrollo de la economía, de la investigación e innovación y para la transformación digital de las Administraciones Públicas y el papel que en estos planes se asigna al turismo cultural y a las instituciones de memoria.

**Palabras clave:** Turismo cultural. Patrimonio cultural digitalizado. Datos abiertos de patrimonio cultural. Datos abiertos. Europeana.

## 1. Introducción (1)

La digitalización del patrimonio cultural de Europa es un proceso al que ha venido a poner orden el proyecto de Europeana. De hecho, uno de los mayores aportes de Europeana es el Europeana Data Model (EDM), que se está aplicando tanto en las bibliotecas como en los archivos o en los museos —dando lugar al acrónimo GLAM, es decir Galleries, Libraries, Archives and Museums. Bajo este acrónimo se agrupan también las pautas comunes con las que las galerías, las bibliotecas, los archivos y los museos están digitalizando su patrimonio y afrontan la transformación a la tecnología Linked Open Data, iniciativa a la que es frecuente referirse como LOD-LAM (acrónimo de Linked Open Data-Libraries, Archives and Museums).

Indudablemente, el patrimonio cultural digitalizado en un entorno Linked Open Data puede jugar un papel fundamental para impulsar todo tipo

## Abstract

The paper reviews the projects developed by Europeana to show the role that digitization, linked open data and cultural heritage can play in strengthening the cultural tourism sector. A series of basic statistical data about tourism and cultural tourism is provided to frame it as an economic sector. The fundamental strategies of the European institutions on cultural tourism are reviewed, as well as the preponderant role they assign to Europeana. These strategies have been disrupted and accelerated in response to the economic debacle caused by the COVID-19 pandemic. The main strategic plans of the Government of Spain for the development of the economy, research and innovation and for the digital transformation of Public Administrations are mentioned, having in mind the role that these plans assign to cultural tourism and memory institutions.

**Keywords:** Cultural tourism. Digital cultural heritage. Cultural heritage. Linked Open Data. Europeana.

de acciones en distintos ámbitos, como por ejemplo, las industrias culturales o creativas, la educación y el turismo.

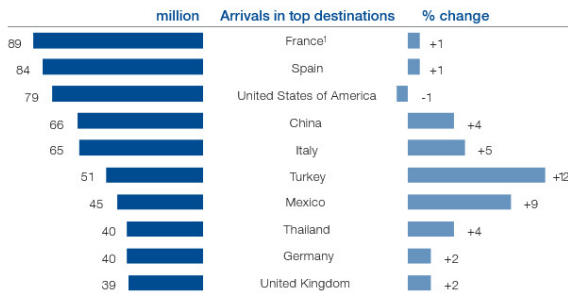
## 2. El turismo cultural

Este artículo está centrado en Europa y en Europeana, como epítome de las bibliotecas virtuales europeas. Sin embargo, dado que la Digital Public Library of America (DPLA) es un proyecto, si no gemelo por lo menos mellizo —de hecho su esquema de metadatos es casi idéntico— siempre es interesante comparar, aunque sea a vista de pájaro, cualquier materia en ambos proyectos. Así, más allá de las similitudes en el modelo de datos, ambas iniciativas difieren especialmente en la forma de afrontar su relación con la educación o con el turismo. Pues bien, la DPLA da muy poca importancia al turismo, lo cual es llamativo porque según la Organización Mundial del Turismo, Estados Unidos es el tercer destino más

apreciado por los turistas a nivel global, por detrás de España y Francia (World Tourism Organization, 2021).

También hay que recordar que según la UNESCO España es el tercer país del mundo, después de China e Italia, con mayor número de bienes inscritos en la lista del Patrimonio de la Humanidad (UNESCO, World Heritage List Statistics).

The world's top 10 destinations receive 40% of global arrivals



Top 10 destinations by international tourist arrivals, 2019\*

<sup>1</sup> Data for France corresponds to 2018. Source: World Tourism Organization (UNWTO). Data as of November 2020. \* Provisional data.

Figura 1. Principales destinos turísticos (Organización Mundial del Turismo, 2020)

En el marco de la Unión Europea, y según las Estadísticas sobre turismo de Eurostat (departamento estadístico de la Unión Europea), el país más visitado, por el número de pernoctaciones, es España, y es también el país con mayores ingresos turísticos netos, 46.300 millones de euros en 2018 (Eurostat, 2020a). Eurostat no proporciona en estas estadísticas qué porcentaje de la actividad turística se refiere al turismo cultural.

La Comisión Europea proporciona cifras muy significativas sobre el turismo en Europa: el 40 % de los turistas de la UE escogen sus destinos por razones culturales y por los sitios de patrimonio cultural. El 62 % de los europeos hacen al menos un viaje de placer al año y la mayoría tiene como destino la propia Europa; el gasto en turismo durante los veranos es de 190 billones de euros; y finalmente, el 10 % del producto interior bruto europeo procede del turismo.

En España, el *Anuario de Estadísticas Culturales* del Ministerio de Cultura y Deporte, publicado en noviembre de 2020 sobre los datos de 2019, indica que el 17 % del total de viajes realizados por los residentes en España fueron iniciados principalmente por motivos culturales, mientras que el 19,8 % de las entradas de turistas internacionales (14,5 millones de viajes) fue principalmente por motivos culturales (España. Ministerio de Cultura y Deporte, 2020).

Eurostat (2020b) ofrece un análisis algo más detallado en *Culture statistics — 2019 edition* como es el *Use of information and communication technology for cultural purposes*. El panorama general que reflejan las estadísticas de cultura en 2019 es que 90 % de los hogares de la UE disponen de acceso a Internet. Holanda, Islandia y Noruega alcanzan el nivel más alto (98 %), cerrando la lista, Grecia (80 %), Montenegro (80 %) y Bulgaria (79 %). España se sitúa en el primer tercio de la tabla con un 95 %.

En cuanto al uso de Internet para fines culturales (2), el análisis parte de una encuesta realizada sobre la población comprendida entre los 16 y los 74 años sobre un conjunto de actividades culturales prefijadas. Para 2019 las actividades mayoritarias son ver televisión o vídeos (70 %), escuchar música (59 %) y jugar o descargar juegos (33 %). El análisis ofrece también datos agregados según franjas de edad, nivel socioeconómico y sexo. (Eurostat 2020c)

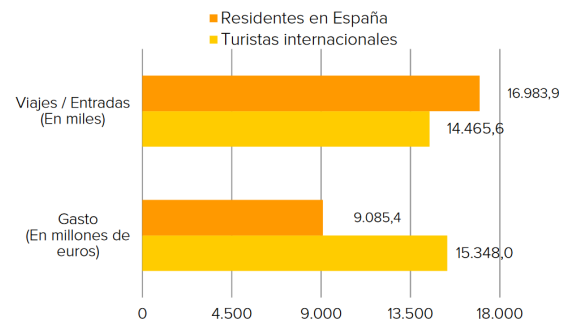


Figura 2. Viajes por motivos culturales y gasto asociado (*Anuario de Estadísticas Culturales, 2020, p. 33*)

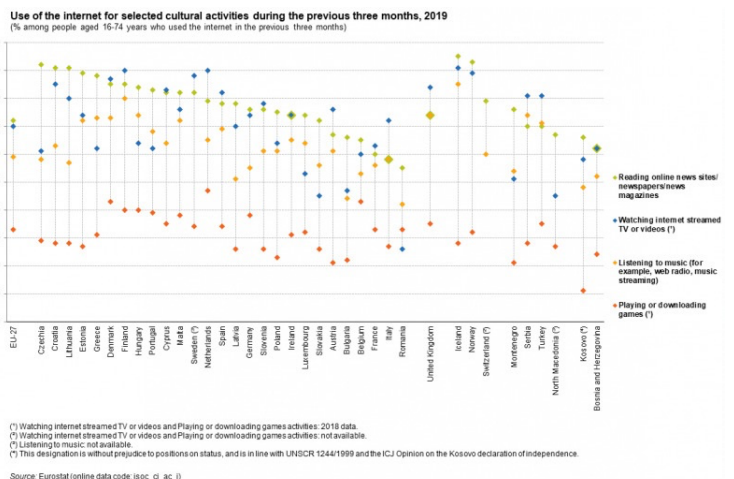


Figura 3. Uso de internet para actividades culturales (Eurostats, *Culture statistics, 2019 edition*)

Aunque los datos estadísticos que hemos reseñado proporcionan un panorama general sobre el

turismo cultural o sobre el uso de Internet con fines culturales, no se recogen datos, ni puede dilucidarse de forma indirecta, el grado de influencia de bibliotecas, archivos y museos digitales en el turismo cultural. Esto nos lleva a una cuestión básica que es cómo podemos valorar, y medir, las áreas de influencia de las bibliotecas, archivos y museos digitales y de sus servicios. Y preguntarnos si es posible basar estos servicios en mediciones más o menos objetivas o, por el contrario, primero es necesario desarrollar los servicios para poder medir su impacto.

Como veremos más adelante, la pandemia ha obligado a las autoridades europeas a fortalecer el turismo cultural por medio del patrimonio digital y a acelerar la creación de servicios destinados a este sector del turismo cultural, basándose más en su urgente necesidad que en sistemas de evaluación de la gestión.

### 3. Europeana: modelo de negocio y análisis de impacto

En 2015, en el marco de la Europeana Strategy 2015-2020 (Europeana, 2017), Europeana lanzó el *Europeana's Impact Assessment Framework*. Su objetivo era medir su impacto y comprobar el grado de cumplimiento de los objetivos declarados en su estrategia. En 2016 Simon Tanner actualizó y ajustó este marco de trabajo e hizo algunas recomendaciones para ponerlo en práctica en el trabajo diario (Tanner, 2016).

Los resultados se presentaron en el documento *MS26: Recommendation report on business model, impact and performance indicators* para el que se utilizó como referencia el texto del citado Simon Tanner *Measuring the Impact of Digital Resources: The Balanced Value Impact Model* (Tanner, 2012). Este es un modelo teórico-práctico sobre cómo establecer las medidas e indicadores para evaluar el impacto de los recursos digitales. Registra también algunos análisis concretos de impacto económico de archivos, bibliotecas y museos, entre los que citaremos el informe final realizado por Archives, Libraries and Museums Alliance UK: *Economic Impact Toolkits for Archives, Libraries and Museums: Final Report* (ALMAUK, 2010). Para trasladar esta capacidad de análisis a las instituciones de memoria, en 2017 Europeana diseñó un conjunto de herramientas y procedimientos, el *Impact Playbook*, en el que ha participado la Impact Assessment Task Force 2.0 de Europeana (Bley, 2018).

Estos documentos no contienen instrucciones o recetas para realizar mediciones o para elaborar indicadores, directamente utilizables por los responsables de bibliotecas, archivos y museos. Además, los análisis de impacto económico que

recogen se han realizado por consultoras especializadas. Se trata más bien de un conjunto de procedimientos para iniciar y configurar el proceso institucional de analizar el impacto de bibliotecas, archivos y museos, más allá de las estadísticas de visitas y visitantes a los respectivos sitios web.

Son también una muestra de la necesidad de justificar a los administradores, que controlan el reparto del presupuesto, que las instituciones de memoria son generadoras de ingresos y por tanto merecedoras de la inversión de fondos públicos. Por debajo permanece la pregunta, que quizá solo tenga una respuesta política, de si en caso de concluir en un balance negativo los archivos, bibliotecas y museos deberían dejar de proporcionar algunos de sus servicios.

En la introducción del *Europeana's Impact Assessment Framework* se dice literalmente:

In the for-profit world, the assessment of a good return on investment is relatively straightforward - how much did our income outweigh the costs, commonly measured in Earnings Before Interest, Taxes, Depreciation and Amortization or EBITDA. This is much harder to do in the not-for profit world where money is an input, but not the right currency for measuring output.

[En el mundo con fines de lucro, la evaluación de un buen retorno de la inversión es relativamente sencilla: en cuánto superaron nuestros ingresos a los costes, comúnmente medidos en Ganancias antes de Intereses, Impuestos, Depreciación y Amortización (EBITDA, por sus siglas en inglés). Esto es mucho más difícil de hacer en el mundo sin fines de lucro, donde el dinero es un insumo, pero no la moneda adecuada para medir la producción.]



Figura 4. Uso de internet para actividades culturales (Eurostats, Culture statistics, 2019 edition)

Sea como sea, la *Europeana Strategy 2015-2020* perfiló e hizo público el modelo de negocio de Europeana, modelo que tuvo su precedente en la propia solicitud de continuación de financiación

bajo los términos de la Connecting Europe Facility, a la que se incorporó la valoración socioeconómica realizada ex profeso por la agencia SEO Economic Research (Europeana, 2017):

The model is based on the principle that cultural heritage should be freely available to end-users and that Europeana needs to develop services that generate income from three different sources: the cultural sector and the Member States that are part of our Network (service fee), the Creative Industry (revenue) and the European Commission (Connecting Europe Facility and project funding from other programmes such as Horizon 2020 and Creative Europe).

[El modelo se basa en el principio de que el patrimonio cultural debe estar disponible gratuitamente para los usuarios finales y que Europeana necesita desarrollar servicios que generen ingresos de tres fuentes diferentes: el sector cultural y los Estados miembros que forman parte de nuestra red (tarifa de servicio), la Industria Creativa (ingresos) y la Comisión Europea (Mecanismo Conectar Europa y financiación de proyectos de otros programas como Horizonte 2020 y Europa Creativa).]

Europeana ha ido construyendo ese modelo de negocio mediante un servicio central formado por los metadatos, el contenido y la tecnología disponible para las instituciones miembro; una capa de acceso a los datos en la cual es primordial el uso de Linked Open Data; y una tercera capa de servicios destinada a tres grupos principales de usuarios, los usuarios finales, los creativos y los profesionales.

En concreto, Europeana identificó distintos grupos de usuarios para los cuales la plataforma ofrecía grandes ventajas: las galerías, bibliotecas, archivos y museos; el público en general; los turistas; las industrias creativas; y las instituciones educativas y los investigadores.

En cuanto al turismo cultural, la idea subyacente en la *Europeana Strategy 2015-2020* es que mejorar el acceso a la información sobre arte, cultura y patrimonio aumentará el atractivo de Europa como foco turístico, especialmente para las regiones menos conocidas. La medida de estos efectos se podría calcular mediante una estimación del porcentaje de crecimiento del turismo y a través de una estimación del incremento en la duración de las pernoctaciones y gastos de los turistas.

Se puede considerar que estas son las estimaciones que mejor medirían el impacto de Europeana en el turismo cultural, pero es también patente la dificultad de obtener estos datos, máxime si Eurostat (es decir, los Estados miembro) no incluye a Europeana como uno de los usos de Internet con fines culturales. Por otra parte, se observa un contraste entre el modelo de negocio de Europeana, y las razones que aducen las instituciones de memoria europeas, uno de sus grupos

de usuarios principales, para proporcionar el acceso digital a sus colecciones.

Estas razones se prefijaron en el diseño de la última encuesta realizada por el *ENUMÉRATE Observatory* (3) en 2017 con los siguientes resultados:

Uso	Puntuación
Uso académico	8,5
Uso educativo	8,5
Reducir el uso de los originales	7,6
Disfrute personal	6,3
Uso y reutilización creativos	5,3
Uso ideológico, religioso o conmemorativo	4,7

*Tabla 1. Razones de las instituciones de memoria para proporcionar acceso digital a sus colecciones en una escala de 0 a 10*

Siendo el objetivo del ENUMERATE Observatory 'proporcionar una base fiable de datos estadísticos sobre digitalización, preservación digital y acceso en línea al patrimonio cultural de Europa' (Agenjo y Campillejo, 2015) se aprecia una cierta descoordinación con la estrategia de Europeana. Parecería lógico que a la hora de diseñar la encuesta se hubieran incluido algunos de los extremos de interés para Europeana como es el turismo cultural, máxime habiéndose realizado la encuesta en el periodo en el que estaba vigente la *Europeana Strategy 2015-2020*.

No hay ninguna duda de que el desarrollo de servicios para el turismo cultural en Europeana es un ejemplo de cómo la plataforma ha ido evolucionando y, también, mostrando a las instituciones de memoria la forma en la que construir nuevos servicios que no estaban originalmente en la intención de los proveedores de datos.

Quizá, si tuviéramos que resumir la influencia de Europeana podríamos centrarla en la consulta multilingüe; la creación de un entorno Linked Open Data; y el desarrollo de servicios de difusión, exploración y explotación de las colecciones digitales, entre las que se encuentra el turismo cultural.

#### 4. Europeana y el turismo cultural

En definitiva, para Europeana es muy importante hacer llegar al usuario de la red, que es un turista en potencia, unos mensajes basados en los contenidos digitalizados, y dado que Europeana tiene como una de sus misiones, quizá la más



importante de todas, preservar y difundir el patrimonio, es evidente la conveniencia de establecer una línea estratégica con ese fin.

Hay que decir de entrada que la *Europeana strategy 2020-2025* (Europeana, 2020) no contempla una línea específica para el turismo cultural, aunque es muy posible que tenga que replantear toda su estrategia en un escenario post-COVID-19. De hecho, la Comisión Europea ha acelerado la creación de servicios de apoyo al turismo cultural con independencia de la estrategia de Europeana. Desde luego, bien puede entenderse que los servicios de turismo cultural están contenidos en ella ya que el objetivo fundamental de Europeana, siguiendo la estrategia política de la Comisión Europea, es apoyar la transformación digital de Europa desde el sector del patrimonio cultural.

Entre las prioridades que se ha marcado Europeana están las de mejorar las capacidades de la infraestructura, mejorar la calidad de los datos, y apoyar la transformación digital de las instituciones de memoria y de los sistemas nacionales de agregación.

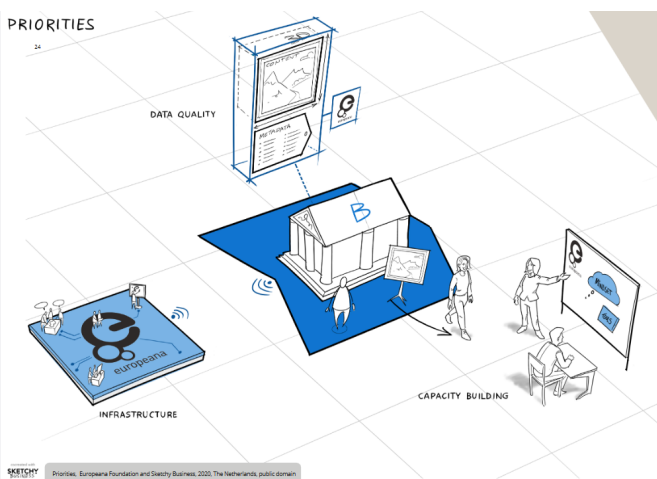


Figura 5. Prioridades  
(Europeana strategy 2020-2025)

En *Enhancing usability of European Digital Cultural Library using web architectures and deep learning* (Machidon et al., 2020) se señala particularmente el problema de la precisión de las búsquedas en Europeana, lo que afecta también a sus API, motivada por la existencia de metadatos de baja calidad, para cuya subsanación proponen un método basado en técnicas inteligentes y de aprendizaje profundo para la interpretación, clasificación y refinado de los resultados en un motor de búsqueda amigable.

Una de las estrategias más importantes de Europeana es proporcionar datos reutilizables a terceros, lo que incluye las instituciones y empresas

dedicadas al turismo. De hecho, el planteamiento de la estrategia de *Europeana para 2020-2025*, 'Prioridad 2 Mejorar la calidad de los datos', es muy similar a la propuesta de Machidon:

Europeana invertirá recursos en actividades relacionadas con la mejora de los metadatos y del contenido en colaboración con agregadores y proveedores de datos. Utilizará nuevas tecnologías como algoritmos de aprendizaje automático para enriquecer automáticamente o semiautomáticamente los metadatos de una manera más rápida y de forma escalable.

Pero la COVID-19 ha alterado todos los planes y estrategias y se hace imprescindible consultar la *Comunicación de la Comisión al Parlamento Europeo... sobre el turismo y el transporte en 2020 y en adelante* (Comisión Europea, 2020a), donde se da un análisis de la importancia del turismo para el empleo y para Europa, destacando cuáles son los puntos débiles y fuertes del turismo, y la importancia de la digitalización del patrimonio cultural. Una de las notas del texto proporciona el siguiente resumen:

El turismo cultural, que representa el 40 % del turismo en Europa, está resultando particularmente afectado, ya que se han cancelado la mayoría de las actividades culturales, como ferias y festivales, e instituciones como los museos han cerrado sus puertas (92 %).

En concreto, hay una llamada a la importancia de Europeana para que contribuya al sostenimiento de este sector económico:

A partir de junio, Europeana, la plataforma europea del patrimonio cultural digital, seguirá desarrollando su vertiente de turismo, donde se muestran las joyas culturales y tesoros ocultos de Europa. En la segunda mitad del año se pondrá en marcha, en la aplicación web complementaria Cultural gems, una campaña de ciudadanos embajadores, con el fin de apoyar el turismo de proximidad.



Bruselas, 29.6.2020

COM(2020) 550  
final/2

#### CORRIGENDUM

This document corrects document COM(2020) 550 final of 13.05.2020.

Concerns the Spanish language version.

In point III. (a), the graph on 'Tourism total contribution to GDP 2018' has been replaced.

The text shall read as follows:

COMUNICACIÓN DE LA COMISIÓN AL PARLAMENTO EUROPEO, AL CONSEJO, AL COMITÉ ECONÓMICO Y SOCIAL EUROPEO Y AL COMITÉ DE LAS REGIONES

El turismo y el transporte en 2020 y en adelante

COMUNICACIÓN DE LA COMISIÓN AL PARLAMENTO EUROPEO, AL CONSEJO, AL COMITÉ ECONÓMICO Y SOCIAL EUROPEO Y AL COMITÉ DE LAS REGIONES

El turismo y el transporte en 2020 y en adelante

Figura 6. El turismo y el transporte en 2020 y en adelante (Corrección de 29 de junio de 2020)

En este nuevo marco y con estos antecedentes, hay que mencionar que el 16 de junio de 2020 el blog de Europeana Pro publicó una entrada titulada *Travelling for pleasure: a brief history of tourism* (Pérez, 2020), redactada por Carmen Pérez Rodríguez, representante de Hispana dentro del Ministerio de Cultura y Deportes. Esta entrada contrasta con la exposición en línea titulada *The Human Factor*, que forma parte de la serie *Heritage at Risk*, donde se alerta sobre cómo la urbanización, la inestabilidad política y el turismo masificado presentan graves amenazas para los monumentos y paisajes de Europa. Seguramente, estas graves amenazas no van a ser tenidas en cuenta, al menos en un futuro próximo, dada la enorme crisis que estamos viviendo desde el año 2020.

Por esto mismo es fundamental revisar la política de la Unión Europea denominada *The Europeana platform: Europe's digital cultural collection for responsible, accessible, sustainable and innovative tourism*, donde ya se parte de la crisis económica que está viviendo Europa por el descenso del turismo durante la COVID-19 y se hace hincapié en el papel que debe jugar Europeana para paliar el drástico descenso del número de turistas (European Commission, 2020).

En efecto, cumpliendo con esa línea estratégica de la Unión Europea y como parte de sus iniciativas, incluidas *Europe's Culture Close to you* y *Cultural Gems*, integradas en el contexto de *Re-open EU* (European Union, 2020), Europeana ha creado una sección especial dentro del portal Europeana Pro titulada *Discovering Europe*, destinada a los profesionales del patrimonio cultural, en la que se ofrecen noticias, contenidos e iniciativas relacionadas con este sector, que ya se anunciaba en *The Europeana platform...* como el futuro Tourism Hub.

Es evidente, aunque no se dice públicamente, que es maravilloso conocer el patrimonio a través de plataformas como Europeana, pero que lo que de verdad se intenta es que los turistas se animen y lleven a cabo sus viajes y estancias a los lugares con alta concentración de patrimonio porque al final es lo que de verdad dinamiza la economía.

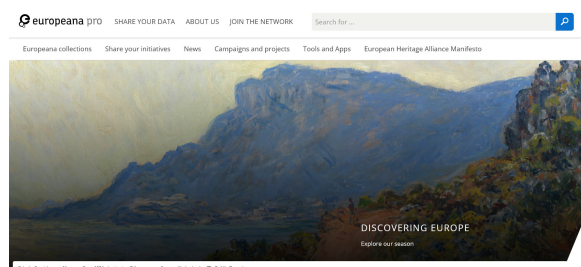


Figura 7. *Discovering Europe* en Europeana Pro

En definitiva, las instituciones europeas han confirmado de facto el papel de Europeana en el sostenimiento del turismo:

Europeana can inspire people to discover beautiful cities and landscapes, historic places and hidden gems across Europe. Our cultural heritage serves as the glue between the past, present and the future helping people to learn about their history.

[Europeana puede inspirar a las personas a descubrir hermosas ciudades y paisajes, lugares históricos y gemas escondidas en toda Europa. Nuestro patrimonio cultural actúa como pegamento entre el pasado, el presente y el futuro, y ayuda a las personas a aprender sobre su historia.]

También, la Comisión Europea ha movilizado a su Centro Común de Investigación para desarrollar la ya citada aplicación para dispositivos móviles *Cultural Gems*, que permite mostrar la riqueza e interés de las ciudades, proporcionando información de patrimonio cultural sobre OpenStreetMap.

Europeana posee una estructura de información codificada según la ontología Europeana Data Model, inmersa completamente en la tecnología Linked Open Data. Sin embargo, todavía no se está realizando la potencialidad de esas herramientas, ni su capacidad para vincular y reutilizar la información. Por ejemplo, sería fácil utilizar los vínculos entre los registros de Europeana con Wikidata y a partir de ahí acceder a la Wikipedia, que es el quinto sitio más utilizado de la Web. Wikidata tiene definida la propiedad 'Europeana entity' (P7704) para identificar personas, lugares y temas en Europeana, pero hasta la fecha sólo tienen esta relación 160.465 ítems de Wikidata.

Por supuesto sería mucho mejor que estos vínculos estuvieran reflejados en los artículos de Wikipedia, pero siempre hay un desfase entre las relaciones de Wikidata con identificadores externos y el 'control de autoridades' de los artículos de Wikipedia. A través de Wikidata y Wikipedia se podría diseñar proyectos específicos de turismo cultural. Y diríamos lo mismo de la utilización de los contenidos de Europeana en Google Knowledge Graph. De hecho, como se sabe, Google Knowledge Graph utiliza profusamente los contenidos de Wikipedia, por lo que sería factible hacer lo mismo con los contenidos de Europeana.

En cualquier caso, las líneas estratégicas y las iniciativas europeas, como *Re-open EU*, se refieren con insistencia a los sitios web que ha lanzado Europeana relacionadas con el turismo cultural, como *Discovering Europe with Europeana Collections*, y a otros como *World Heritage Journeys of the European Union*, *The European Capital of Smart Tourism*, el proyecto financiado por la Unión Europea IMPACTOUR, los *European*

*Heritage Awards/Europa Nostra Awards*, o el *European Heritage Alliance Manifesto*. Como se ha mencionado, *Discovering Europe* ha devenido en el Tourism Hub que se anunciaba en la política de la Comisión *The Europeana platform: Europe's digital cultural collection for responsible, accessible, sustainable and innovative tourism*.

Además, hay que mencionar una serie de proyectos de reutilización del patrimonio por terceros, como son las apps *Greetings from Zagreb*, *VanGoYourself*, y *Virtual 17th-century Flemish inn*, no solamente desde el punto de vista del turismo, sino en general.

Desde luego, es encomiable la rápida respuesta para recoger las distintas experiencias que las instituciones de memoria de todo el mundo han puesto en marcha para sobrellevar la COVID-19, con las que se ha compuesto la página *Creative Approaches and Collaborations* de Europeana Pro, especialmente los tres finales: *Collecting memories of COVID-19 in Sweden* del Nordiska Museet; *COVID-19 in Vienna: A collection project on the city's history* del Vienna Museum; y *These Times* del Museum of Ordinary People. También queremos destacar un proyecto muy interesante, lanzado en estos meses de pandemia, se trata de *My Home is My Museum* cuyo autor son los National Museums Liverpool. En este proyecto se da mucha importancia a cómo los niños pueden ver el patrimonio, y sobre todo a las ideas que un niño puede plasmar al crear una lista de 10 objetos que puede escoger para diseñar su propia exposición. A pesar de su sencillez, incorpora más creatividad que mucho de lo que se ha hecho en este año de 2020.

Seguramente una de las selecciones más completas de portales, tours virtuales, e-learning y colecciones en línea es la recogida en el sitio web del MCN - *Advancing Digital Transformation in Museums*. Y, finalmente, no podemos dejar de mencionar, aunque solo sea de pasada, el esfuerzo que están realizando muchas instituciones culturales y científicas para exponer sus fondos y contenidos de una forma intensa y novedosa en el marco de Google Arts and Culture, así como la competencia que este conjunto de herramientas y app va a plantear, con toda seguridad, a Europeana.

Por nuestra parte nos gustaría mencionar el proyecto Heritage Hubs, porque guarda cierta similitud con el de la ciudad natal de los Beatles y porque en él participa la Fundación Ignacio Larramendi. Este proyecto, financiado por la Unión Europea, ha reunido a colegios de Finlandia, Serbia, Italia y España, y sus resultados se podrán ver próximamente en la Biblioteca Virtual Heritage Hubs. Muy recientemente este proyecto ha sido

galardonado con el premio Europa Nostra Awards de 2021.



Figura 8. Biblioteca Virtual HeritageHubs

## 5. Las líneas estratégicas españolas

Es muy importante mencionar tres estrategias digitales que el Gobierno de España ha publicado en los últimos meses. Por supuesto, hay que tomar con las debidas precauciones todo lo que se está enunciando en esos tres planes porque han sido publicados en plena pandemia (cuando se revisa este artículo todavía España, y en general todo el mundo, está batallando con la tercera ola de la COVID-19).

El primer plan recibe el nombre de *Estrategia Española de Ciencia, Tecnología e Innovación 2021-2027* y fue aprobada en Consejo de Ministros el 8 de septiembre de 2020 (España. Secretaría General de Investigación, 2020). Este plan menciona de pasada al turismo, englobando, y esto es muy importante, en una misma área las industrias culturales y el turismo (dentro de la sublínea Mundo digital, Industria, Espacio y Defensa de la línea Inteligencia Artificial y Robótica).

La *Estrategia Nacional de Inteligencia Artificial (ENIA)* (España. Gobierno, 2020), presentada el 2 de diciembre de 2020, dispone una inversión pública de 600 millones para el periodo 2021-2023 y supone una concreción del *Plan España Digital 2020-2025*, así como es el componente 16 del *Plan de Recuperación, Transformación y Resiliencia* de la economía española. Esta estrategia agrupa de nuevo el turismo y las industrias creativas, haciendo hincapié en que ambas constituyen un campo de cultivo para la Inteligencia Artificial, clasificándolas incluso como el sexto sector con mayor impacto esperado de la Inteligencia Artificial a corto y medio plazo, aunque no siempre queda muy claro si cuando se refiere al sector del turismo está ligado o no con las industrias creativas.

El tercer plan, aprobado el 30 de diciembre de 2020, que lleva el membrete *España Puede* y el subtítulo de *Plan de Recuperación, Transformación y Resiliencia* (España. Presidencia del Gobierno, 2020), enuncia 10 políticas *palanca* de reforma estructural para un crecimiento sostenible e inclusivo. Entre ellas, y con el número 5, se menciona una política para, entre otras cosas, la recuperación del turismo, puesto que como muy bien dice el *Plan*, el turismo juega un papel crucial en la economía de España y además de suponer un 10 % del PIB genera importantes efectos tractors sobre otros sectores. Pero, aunque esta estrategia incluye los planes de digitalización de cuatro sectores estratégicos: salud, automoción, turismo y comercio, además del agroalimentario, no se hace una mención especial al turismo cultural, aunque sí a las industrias culturales y al turismo deportivo (Política 9 Impulso de la industria de la cultura y el deporte).

Por último, el 27 de enero de 2021 se aprobó el *Plan de Digitalización de las Administraciones Públicas 2021-2025* (España. Gobierno, 2020), uno de los tres planes clave del *Plan España Digital 2020-2025* junto con el *Plan Nacional de Competencias Digitales* y el *Plan de Digitalización de Pymes*. Según se anuncia, estos tres planes estratégicos dispondrán de una inversión conjunta de 11.000 millones de euros y su objetivo es reforzar la digitalización para aumentar la productividad y el crecimiento económico, impulsar la creación de empleo de calidad y conquistar los mercados exteriores. Sin embargo, no hay mención alguna al turismo, al turismo cultural, o a la cultura.

## 6. Conclusión

Este artículo, redactado en medio de una pandemia que dura ya más de un año si tenemos en cuenta la aparición de los primeros casos en España, intenta llamar la atención de los actores del patrimonio cultural, especialmente a los bibliotecarios, archiveros y museólogos, sobre el papel que pueden jugar a la hora de impulsar el turismo cultural, basándose en la estructura de información de Europeana, especialmente del uso de la tecnología Linked Open Data.

Sin embargo, a pesar del gran interés de las distintas iniciativas y aplicaciones, desde nuestro punto de vista se debería profundizar en una de las capacidades más importantes de Europeana, su entorno Linked Open Data. Únicamente la nueva app *Cultural Gems* se vincula con Open StreetMap, cuando es palmario que se pueden vincular muchísimos otros conjuntos de información Linked Open Data. Quizá se deba a la ur-

gencia de responder inmediatamente con propuestas relacionadas con el turismo cultural para contrarrestar en algo los efectos de la pandemia.

Se puede también hacer una reflexión sobre la forma en la que la Comisión Europea y las instituciones europeas en general, ante la magnitud de la crisis, han apreciado rápidamente las posibilidades que ofrece Europeana en el sostenimiento del turismo, sin requerir la correspondiente justificación basada en detallados análisis de impacto socioeconómico, ni planes estratégicos adicionales. Además, la Comisión Europea intenta potenciar el turismo cultural aunque pueda chocar con los riesgos de un turismo masificado que se observaban antes de la pandemia.

En cualquier caso, sería una pena que no se aprovecharan las acciones que está impulsando la Unión Europea para fortalecer al turismo cultural, pero el cisne negro que está suponiendo la infección de la COVID-19 está todavía muy presente y seguro que Europeana incluirá en sus planes estratégicos esta posibilidad.

## Notas

- (1) Este artículo se basa en la exposición de los autores en la XXV edición de los Encuentros Internacionales sobre Sistemas de Información y Documentación, que tuvo lugar del 30 de septiembre al 2 de octubre de 2020 en la Biblioteca María Moliner de la Facultad de Filosofía y Letras de la Universidad de Zaragoza. El tema central de estos encuentros fue la relación entre las instituciones de memoria y el turismo cultural. No se le escapará a ningún lector que si este artículo se hubiera redactado antes de la pandemia de la COVID-19, el resultado habría sido muy diferente.
- (2) La tablas 'Number of trips by purpose' de las Estadísticas sobre turismo de Eurostat (Eurostat Data Browser) solo recogen los siguientes motivos: Personal reasons; Professional, Business; Hollydays, Leisure and recreation; Visits to friends and relatives; y Other personal reasons. <https://ec.europa.eu/eurostat/databrowser/view/tin00188/default/table?lang=en>.
- (3) El proyecto ENUMERATE actual tuvo su origen en un proyecto europeo denominado NUMERIC, que continuó en el proyecto ENUMERATE (2011-2014) financiado por el programa ICT de la Comisión Europea, en cuyo consorcio participó DIGIBÍS y los autores de esta comunicación. En 2014 ENUMERATE pasó a formar parte de Europeana v3, transformándose en el ENUMERATE Observatory.

## Referencias

- [Todos los enlaces están comprobados a fecha de 31 de enero de 2021]
- About ENUMERATE. // Europeana Pro, March 24, 2015. <https://pro.europeana.eu/page/about-enumerate>.
- Agenjo, Xavier; Campillejo, Maribel (2015). El observatorio de la digitalización en Europa. // Desafíos y oportunidades de las Ciencias de la Información y la Documentación en la era digital: actas del VII Encuentro Ibérico EDICIC 2015

- (Madrid, 16 y 17 de noviembre de 2015). Madrid: Universidad Complutense de Madrid, 2015. ISBN 978-84-608-3330-7. <https://eprints.ucm.es/34723/>.
- Archives, Libraries, and Museums Alliance UK (2010). Economic Impact Toolkits for Archives, Libraries and Museums: Final Report [2010]. [https://almuk.files.wordpress.com/2010/09/alma-uk-economic-impacts-project-stage-1\\_final-report.pdf](https://almuk.files.wordpress.com/2010/09/alma-uk-economic-impacts-project-stage-1_final-report.pdf).
- Biblioteca Virtual Heritage Hubs. Madrid: Fundación Ignacio Larramendi, 2020. [http://www.larramendi.es/her\\_hub/es/micrositios/inicio.do](http://www.larramendi.es/her_hub/es/micrositios/inicio.do).
- Bley, Jens (2018). What the Europeana Impact Playbook taught a city about revitalising an inner city square. // *Europeana Pro*. November 5, 2018. <https://pro.europeana.eu/post/what-the-europeana-impact-playbook-taught-a-city-about-revitalising-an-inner-city-square>.
- Byrd-McDevitt, Lori (2020). The Ultimate Guide to Virtual Museum Resources, E-Learning, and Online Collections. // *MCN Blog*. March 14, 2020. <https://mcn.edu/a-guide-to-virtual-museum-resources/>.
- Collaborative Approaches and Collaborations // *Europeana Pro*. <https://pro.europeana.eu/page/creative-approaches-and-collaborations>.
- Comisión Europea (2020). Comunicación de la Comisión al Parlamento Europeo, al Consejo, al Comité Económico y Social Europeo y al Comité de las Regiones: El turismo y el transporte en 2020 y en adelante. Bruselas, 13.5.2020, COM(2020) 550 final. <https://eur-lex.europa.eu/legal-content/ES/TXT/HTML/?uri=CELEX:52020DC0550&from=EN>
- Comisión Europea. Estrategia - Prioridades 2019-2024. Una Europa adaptada a la era digital: capacitar a las personas con una nueva generación de tecnologías. [https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age\\_es](https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age_es).
- Cultural gems: Discover cultural and creative places in European cities <https://culturalgems.jrc.ec.europa.eu/>.
- Discovering Europe. // *Europeana Pro*. <https://pro.europeana.eu/page/discovering-europe>.
- ENUMERATE Observatory. // *Europeana Pro*. <https://pro.europeana.eu/page/enumerate>.
- España. Gobierno (2020). Estrategia Nacional de Inteligencia Artificial. <https://www.lamoncloa.gob.es/presidente/actividades/Documents/2020/ENIA2B.pdf>.
- España. Gobierno (2020). Plan España Digital 2025. [https://www.lamoncloa.gob.es/presidente/actividades/Documents/2020/230720-Espa%C3%B1aDigital\\_2025.pdf](https://www.lamoncloa.gob.es/presidente/actividades/Documents/2020/230720-Espa%C3%B1aDigital_2025.pdf).
- España. Ministerio de Cultura y Deporte (2020). Anuario de Estadísticas Culturales 2020. <https://www.culturaydeporte.gob.es/dam/jcr:52801035-cc20-496c-8f36-72d09ec6d533/anuario-de-estadisticas-culturales-2020.pdf>. Véase también el apartado Turismo Cultural del Anuario de Estadísticas Culturales 2020. Principales resultados. Página 33.
- España. Presidencia del Gobierno (2020). España puede: Plan de Recuperación, Transformación y Resiliencia. Madrid, octubre 2020. NIPO: 089-20-023-7. <https://www.lamoncloa.gob.es/presidente/actividades/Paginas/2020/espana-puede.aspx>.
- España. Presidencia de Gobierno (2020), Plan de Digitalización de las Administraciones Públicas 2021-2025: Estrategia en materia de Administración Digital y Servicios Públicos Digitales. <https://www.lamoncloa.gob.es/presidente/actividades/Documents/2021/270121-PlanDigitalizacionAdministracionesOptimizado.pdf>.
- España. Secretaría General de Investigación (2020). Estrategia Española de Ciencia, Tecnología e Innovación 2021-2017. Madrid: Gobierno de España; Ministerio de Ciencia e Innovación, 2020. e-NIPO 83120011X. <https://www.ciencia.gob.es/portal/site/MICINN/menuitem.26172fcf4eb029fa6ec7da6901432ea0/?vgnnextoid=1387571a3db06610VgnVCM1000001d04140aRCRD>.
- Europe's Culture – close to you. <https://ec.europa.eu/culture/news/europes-culture-close-you#:~:text=this%20campaign%20promotes%20the%20re,a%20safe%20and%20sustainable%20way>.
- European Capital of Smart Tourism: an EU initiative to reward innovative and smart tourism in European Cities. <https://smarttourismcapital.eu/>.
- European Commission. European Commission Strategy - Shaping Europe's digital future -Policies: The Europeana platform: Europe's digital cultural collection for responsible, accessible, sustainable and innovative tourism. <https://ec.europa.eu/digital-single-market/en/europeana-platform-europes-digital-cultural-collection-responsible-accessible-sustainable-and>.
- European Heritage Alliance Manifesto. [https://www.europanostr.org/wp-content/uploads/2020/09/20200915-EHA-MANIFESTO-layout\\_high-res.pdf](https://www.europanostr.org/wp-content/uploads/2020/09/20200915-EHA-MANIFESTO-layout_high-res.pdf).
- European Union (2020). Re-open EU [2020]. <https://reopen.europa.eu/es>.
- Europeana (2014). We transform the world with culture: Europeana Strategy 2015-2020. [2014] [https://pro.europeana.eu/files/Europeana\\_Professional/Publications/Europeana%20Strategy%202020.pdf](https://pro.europeana.eu/files/Europeana_Professional/Publications/Europeana%20Strategy%202020.pdf). Véase también "Europeana 2020 Strategic update". [2017] <https://strategy2020.europeana.eu/update/>.
- Europeana (2017). Europeana DSI 2– Access to Digital Resources of European Heritage. D4.4. [https://pro.europeana.eu/files/Europeana\\_Professional/Projects/Project\\_list/Europeana\\_DSI-2/Deliverables/d4.3%20DSI-2%20Final%20Project%20Report.pdf](https://pro.europeana.eu/files/Europeana_Professional/Projects/Project_list/Europeana_DSI-2/Deliverables/d4.3%20DSI-2%20Final%20Project%20Report.pdf)
- Europeana (2017). Report on ENUMERATE Core Survey 4. [http://pro.europeana.eu/files/Europeana\\_Professional/Projects/Project\\_list/ENUMERATE/deliverables/DSI-2\\_Deliverable%20D4.4\\_Europeana\\_Report%20on%20ENUMERATE%20Core%20Survey%204.pdf](http://pro.europeana.eu/files/Europeana_Professional/Projects/Project_list/ENUMERATE/deliverables/DSI-2_Deliverable%20D4.4_Europeana_Report%20on%20ENUMERATE%20Core%20Survey%204.pdf).
- Europeana (2020). Strategy 2020-2025: Empowering Digital Change. 2020. <https://pro.europeana.eu/page/strategy-2020-2025-summary>.
- Eurostat (2020). Culture statistics - 2019 Edition. [https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Culture\\_statistics\\_-\\_introduction](https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Culture_statistics_-_introduction).
- Eurostat (2020). Culture statistics - use of ICT for cultural purposes. Data extracted in June 2019. [https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Culture\\_statistics\\_-\\_use\\_of\\_ICT\\_for\\_cultural\\_purposes](https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Culture_statistics_-_use_of_ICT_for_cultural_purposes).
- Eurostat (2020). Estadísticas sobre turismo. Datos de enero de 2020. [https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Tourism\\_statistics/es](https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Tourism_statistics/es).
- Eurostat (2021). Households - level of internet access (2011-2020). Datos actualizados a 26/01/2021. [http://ec.europa.eu/eurostat/product?code=isoc\\_ci\\_in\\_h&language=en&mode=view](http://ec.europa.eu/eurostat/product?code=isoc_ci_in_h&language=en&mode=view).
- Google Arts & Culture. <https://artsandculture.google.com/>.
- Greetings from Zagreb. <https://pro.europeana.eu/data/greetings-from-zagreb>.
- Heritage at Risk: Protecting and Preserving Endangered Cultural Heritage // *Europeana Pro*. <https://www.europeana.eu/en/exhibitions/heritage-at-risk>.
- Heritage Hubs. <https://heritagehubs.eu/es/>.
- The Human Factor // *Europeana Pro*. <https://www.europeana.eu/en/exhibitions/heritage-at-risk/the-human-factor>.
- The Impact Assessment Task Force 2.0 // *Europeana Pro*, January 25, 2018. <https://pro.europeana.eu/project/impact-assessment-2-0-task-force>.
- Impact: How do we know we are making a difference? // *Europeana Pro* <https://pro.europeana.eu/page/impact>.

- Machidon, Octavian; Stoica, Dragoş; y Tavčar, Aleš (2020). Enhancing the Usability of European Digital Cultural Library Using Web Architectures and Deep Learning. // Cultural and Tourism Innovation in the Digital Era: Sixth International IACuDiT Conference, Athens 2019. Springer, 2020, pp.201-207. 2020. [https://doi.org/10.1007/978-3-030-36342-0\\_16](https://doi.org/10.1007/978-3-030-36342-0_16).
- MCN - Advancing Digital Transformation in Museums. <https://mcn.edu/>.
- Museum of Ordinary People. <https://www.museumofordinarypeople.com/>.
- National Museums Liverpool. <https://www.liverpoolmuseums.org.uk/my-home-my-museum>.
- Nordiska Museet. Coronaviruset i Sverige – berätta för framtiden. <https://minnen.se/tema/corona>.
- Pérez Rodriguez, Carmen (2020). Travelling for pleasure: a brief history of tourism // Europeana Pro, June 16, 2020. <https://blog.europeana.eu/2020/06/travelling-for-pleasure-a-brief-history-of-tourism/>.
- Tanner, Simon (2012). Measuring the Impact of Digital Resources: The Balanced Value Impact Model. London: King's College, 2012. [https://www.kdl.kcl.ac.uk/fileadmin/documents/pubs/BalancedValueImpactModel\\_SimonTanner\\_October2012.pdf](https://www.kdl.kcl.ac.uk/fileadmin/documents/pubs/BalancedValueImpactModel_SimonTanner_October2012.pdf).
- Tanner, Simon (2016). Updating the Europeana Impact Assessment Framework: Recommendations on implementing the Europeana Impact Assessment Framework. [https://pro.europeana.eu/files/Europeana\\_Professional/Projects/Project\\_list/Europeana\\_DSI/Milestones/europeana-dsi-ms26-recommendation-report-on-business-model-impact-and-performance-indicators-2016.pdf](https://pro.europeana.eu/files/Europeana_Professional/Projects/Project_list/Europeana_DSI/Milestones/europeana-dsi-ms26-recommendation-report-on-business-model-impact-and-performance-indicators-2016.pdf).
- UNESCO. World Heritage List Statistics. Number of World Heritage properties inscribed by each State Party. <https://whc.unesco.org/en/list/stat>.
- UNESCO; National Geographic. World Heritage Journeys. Europe. Co-funded by the European Union. <https://visit-worldheritage.com/en/eu>.
- VanGoYourself. <https://pro.europeana.eu/data/vangoyourself>.
- Virtual 17th-century Flemish inn. <https://pro.europeana.eu/data/virtual-17th-century-flemish-inn>.
- Wien Museum. Corona inWien: ein Sammlungsprojekt zur Stadtgeschichte <https://www.wienmuseum.at/de/corona-sammlungprojekt>.
- World Tourism Organization (2021). International Tourism Highlights, 2020 Edition. Madrid: UNWTO, 2020. DOI: <https://doi.org/10.18111/9789284422456>.

---

Enviado: 2021-01-31. Segunda versión: 2021-06-06.  
Aceptado: 2021-06-17.

---

---

# Enriquecimiento automático de portales culturales mediante modelos de organización del conocimiento

*Automatic enrichment of cultural portals with knowledge organization systems*

---

**Dayany DÍAZ-CORONA (1), Javier LACASTA (2), Javier NOGUERAS-ISO (3)**

(1) Instituto de Investigación en Ingeniería de Aragón (I3A), Universidad de Zaragoza, C/ María de Luna, 1, 50018 Zaragoza, España, dayanydc@unizar.es (2) jlacasta@unizar.es (3) jnog@unizar.es

## Resumen

Durante las últimas décadas se han ido creando numerosos portales web para diseminar el patrimonio cultural. La mayoría de estos portales se crearon en tiempos de la web sintáctica generando páginas HTML con texto plano indexable por buscadores, pero sin metadatos añadidos y anotaciones de conceptos pertenecientes a modelos de organización del conocimiento que facilitarían la labor de buscadores temáticos especializados. Este artículo propone un método para recomendar los modelos de organización del conocimiento que mejor se ajusten a los contenidos de un portal web, y utilizar esos modelos para anotar semánticamente los contenidos. Para verificar la viabilidad del método propuesto se ha aplicado en el enriquecimiento de un portal creado a mediados de los años 90 y que aloja un catálogo virtual de las obras del pintor Goya. Gracias al método propuesto, se ha recomendado el modelo de organización del conocimiento denominado Lista de Encabezamientos de Materias para las Bibliotecas Públicas por su cercanía con el contenido del portal. Además, se han conseguido anotar semánticamente dos tercios de las páginas en castellano del portal con conceptos de este modelo. Aunque la exactitud de los emparejamientos entre las entidades detectadas en el texto y los conceptos del modelo no es perfecta, la anotación realizada constituye una buena base para que los administradores del portal puedan refinar posteriormente esta anotación.

**Palabras clave:** Modelos de organización del conocimiento. Sistemas de organización del conocimiento. SKOS. Web semántica. Metadatos. Sistemas de recomendación.

## 1. Introducción

Desde mediados de los años 90 se ha fomentado la digitalización del patrimonio cultural creándose numerosos portales web para difundir ese patrimonio a todos los públicos. La mayoría de estos portales se crearon con tecnologías pertenecientes a la época de la web sintáctica y contienen páginas HTML con texto plano indexable por buscadores. Pero para que estos portales sean indexables por motores de búsqueda y agregadores

## Abstract

During the last decades, numerous web portals have been launched to disseminate the cultural heritage. Most of these portals were developed with technologies from the syntactic web era, i.e. containing HTML pages with plain text that can be indexed by search engines, but without additional metadata and annotations of concepts belonging to knowledge organization systems that would facilitate the task of thematic specialized search engines. This paper proposes a method for recommending the knowledge organization systems that are better adjusted for the contents of a web portal and the use of these systems for the semantic annotation of the contents. To check the feasibility of the proposed method, we have applied it to the enrichment of a web portal created in the nineties that hosts a virtual catalogue of the works performed by the painter Goya. Thanks to the proposed method, we have been able to recommend knowledge organization system titled List of Subject Headings for Public Libraries because of its closeness with the portal content. In addition, two thirds of the web pages in Spanish were annotated with concepts belonging to this model. Although the accuracy of the mapping between the recognized entities in the text and the concepts of the model is not perfect, it constitutes a good base to allow web portal administrators to refine later this annotation.

**Keywords:** Knowledge organization systems. SKOS. Semantic web. Metadata. Recommendation systems.

especializados y que sus contenidos sean fácilmente accesibles por el público, es esencial anotar semánticamente sus contenidos con palabras clave temáticas extraídas de modelos de organización del conocimiento de alta calidad. Estos modelos de organización del conocimiento son vocabularios especializados que incluyen tesauros y cuadros de clasificación, entre otros modelos, y que al utilizarlos de forma consensuada minimizan lo máximo posible los problemas derivados de la heterogeneidad del lenguaje como la

polisemia o la sinonimia (Lacasta et al., 2010; Whaley et al., 2020).

En los últimos años han surgido distintas aproximaciones para analizar la calidad de modelos de organización del conocimiento (Suominen y Mader, 2014; Albertoni et al., 2016; Lacasta et al., 2016). Este análisis de calidad permite que los creadores de contenido puedan seleccionar un modelo en función de su calidad. Sin embargo, este análisis no tiene en cuenta si existe un emparejamiento adecuado entre los contenidos que se quieren anotar temáticamente y los conceptos cubiertos por el modelo de organización del conocimiento. Algunos autores como da Silva Lemos y Souza (2020) proponen una serie de criterios para poder comparar las ventajas y desventajas del uso de distintos sistemas de organización del conocimiento, pero no es un procedimiento automatizado.

El objetivo de esta contribución es proponer un método que permita recomendar de forma automática los modelos de organización del conocimiento que mejor cubran los contenidos de un portal web que se quieren enriquecer semánticamente y facilitar así la capacidad de búsqueda sobre esos contenidos. Es decir, partiendo de un conjunto de modelos de organización del conocimiento con calidad contrastable, se pretende elegir aquel que sea más adecuado al contenido que se quiere anotar.

En la literatura existen varias propuestas de anotadores semánticos de contenido textual de carácter general como DBPedia Spotlight (Daiber et al., 2013) o RDFSFace (Khalili y Auer, 2015). Estas herramientas permiten detectar secuencias de palabras correspondientes a entidades nombradas (o simplemente entidades) que hacen referencia a materias, lugares u organizaciones y las enlazan con conceptos de ontologías o grafos de conocimiento como DBPedia (en el caso de DBPedia Spotlight) u otras fuentes consultables a través de APIs como Swoogle o Síndice. También han surgido propuestas similares de anotación en contextos más específicos como el de la medicina (Thessen y Parr, 2014; Tchechmedjiev et al., 2018) o el sector multimedia (Rodríguez Rocha et al., 2015) focalizadas en el uso de ontologías más específicas o añadiendo técnicas más expertas en el reconocimiento de entidades en estos contextos.

A diferencia de las soluciones existentes, la novedad del trabajo que se propone en este artículo es establecer un flujo de trabajo que permita flexibilizar las ontologías elegibles para la anotación. El método propuesto en este artículo parte de un trabajo previo de los autores (Díaz-Corona et al., 2019) donde se estableció un proceso para

caracterizar los modelos de organización del conocimiento más utilizados en el dominio cultural y que además incluye un análisis de la completitud, consistencia y exactitud de los modelos compatibles con SKOS (Miles y Brickley, 2005), una iniciativa de W3C para representar modelos de organización del conocimiento utilizando un vocabulario RDF. A continuación, para identificar si los modelos con una calidad razonable cubren el contenido que se quiere anotar se evalúa su similitud en el espacio vectorial. Y finalmente, para anotar el contenido con conceptos concretos el método se apoya en software existente de reconocimiento de entidades.

Como banco de pruebas para estudiar la viabilidad de este método de recomendación se ha utilizado el portal InfoGoya (Universidad de Zaragoza, 1996), un catálogo virtual de las obras del pintor Goya creado a mediados de los 90. Es un ejemplo de un portal de la web sintáctica que se puede enriquecer con micro-metadatos para facilitar su semantización y mejorar su indexación.

El resto del artículo se estructura de la siguiente forma. En la sección 2 se explica el método propuesto. En la sección 3 se muestra el resultado de aplicar el método sobre el portal mencionado anteriormente. Finalmente, la sección 4 presenta las conclusiones finales y las líneas de trabajo futuro.

## 2. Método

La figura 1 muestra el procedimiento propuesto en este trabajo para el enriquecimiento semántico de portales.

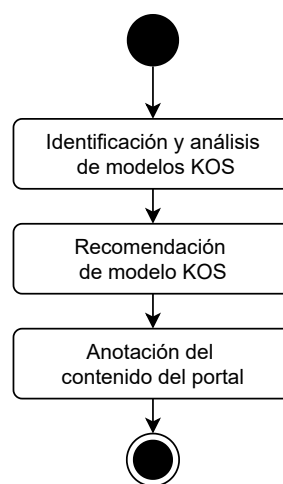


Figura 1. Flujo de trabajo para el enriquecimiento de portales semánticos

El primer paso es la identificación y análisis de potenciales modelos de organización del conocimiento, denominados a partir de este momento



por su acrónimo en inglés KOS (*Knowledge Organization Systems*). Partiendo de la selección previa de modelos KOS, el segundo paso está orientado a recomendar el modelo KOS que resulte más apropiado. El tercer paso es anotar el contenido del portal con el modelo KOS recomendado. En las siguientes secciones se explican los pasos de este procedimiento.

### 2.1. Identificación y análisis de modelos de organización del conocimiento

Como se ha mencionado en la introducción, para la identificación y análisis de los modelos KOS se parte de un trabajo previo de los mismos autores de este artículo que analizaba los modelos KOS utilizados en repositorios de datos enlazados del dominio cultural (Díaz-Corona et al., 2019). Las fases contempladas en el proceso planteado en este trabajo previo se muestran en la Figura 2.

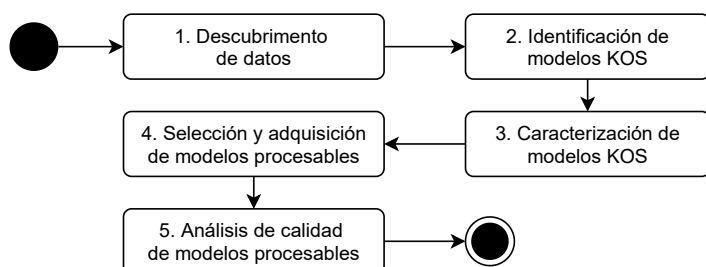


Figura 2. Fases de la metodología para la identificación y análisis de modelos de organización del conocimiento

La primera fase consiste en identificar las fuentes de datos enlazados relevantes para el dominio deseado, así como los métodos de acceso. Es un proceso principalmente manual porque no hay un catálogo único centralizado de fuentes de datos enlazados o un único sistema de búsqueda apropiado para buscar fuentes de datos. Por tanto, es necesario revisar tanto posibles estados del arte referenciando a fuentes de datos como utilizar buscadores de datos (ej: Google Dataset), o incluso buscadores genéricos.

```

<ore:Proxy rdf:about="http://data.europeana.eu/proxy/provider/2059210/data_sounds_http__imslp_org_wiki_Fa__la_nonna_ninno_mio__Sarrìa__Enrico__">
  <dc:creator rdf:resource="http://imslp.org/wiki/Category:Sarrìa,_Enrico"></dc:creator>
  <dc:title>Fa' la nonna ninno mio (Sarrìa, Enrico)</dc:title>
  <dc:subject rdf:resource="http://data.europeana.eu/concept/soundgenres/Music/Western_classical_music"></dc:subject>
  <dc:subject xml:lang="en">Romantic</dc:subject>
  <ore:proxyIn rdf:resource="http://data.europeana.eu/aggregation/provider/2059210/data_sounds_http__imslp_org_wiki_Fa__la_nonna_ninno_mio__Sarrìa__Enrico__"/>
</ore:Proxy>
  
```

Figura 3. Fragmento de los metadatos de un recurso cultural enlazando a un KOS de Europeana

La segunda fase consiste en estudiar el modelo de metadatos de cada fuente con el objetivo de identificar las propiedades de anotación temática que enlazan a conceptos pertenecientes a modelos KOS y extraer un listado de todos los modelos KOS referenciados. Por ejemplo, en la Figura 3 de muestra un fragmento de los metadatos de un recurso sonoro anotado con un concepto del tesoro *Europeana Sounds* a través de la etiqueta *dc:subject*.

La tercera fase es el estudio y caracterización de los modelos identificados en la segunda fase. Como resultado de esta fase se anota cada modelo KOS con los siguientes elementos, entre otros descriptores: información descriptiva como el nombre, su acrónimo y editor; información sobre su relevancia mediante su nº de usos y nº de proveedores de recursos culturales referenciando a estos modelos; tipo de clasificación (temática, de autoridad o de lugar) para la que se utiliza el KOS; y los detalles sobre el mecanismo de acceso y formato de descarga (por ejemplo, SKOS, RDF, OWL, etc.).

La cuarta fase consiste en filtrar y acceder a aquellos modelos KOS utilizados para clasificaciones temáticas de recurso, que hayan sido utilizados por un número relevante de proveedores (dos o más) y estén disponibles en un formato compatible con SKOS.

La última fase está dedicada a analizar en detalle la calidad de los modelos KOS procesables que se descargaron en la fase previa, y que son susceptibles de ser reutilizables para el enriquecimiento semántico de nuevos contenidos. Este análisis, propuesto por Lacasta et al. (2016), está basado en la norma internacional ISO 25964 para tesauros y plantea 14 métricas agrupadas en 3 categorías: consistencia, planteando métricas que chequean el contenido de las etiquetas de los conceptos incluidos en el KOS para asegurar, por ejemplo, un uso uniforme de singulares/plurales o mayúsculas; completitud, agrupando métricas que verifican la existencia de propiedades obligatorias como etiquetas preferidas; y exactitud, incluyendo métricas que chequean, entre otros aspectos, la corrección y exactitud de relaciones jerárquicas entre conceptos.

### 2.2. Recomendación del modelo KOS

El paso anterior del procedimiento ha conseguido identificar modelos KOS con calidad contrastable y un formato apto (compatibilidad con SKOS) para anotar semánticamente contenidos, pero no hemos verificado si estos vocabularios son adecuados temáticamente para nuestros contenidos.

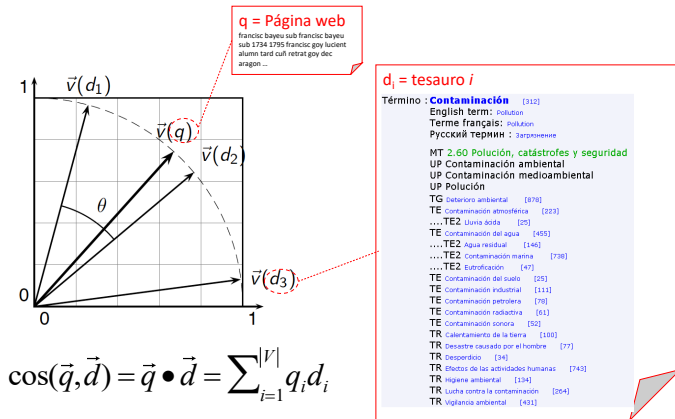


Figura 4. Adaptación del modelo vectorial para el ranking de modelos KOS

#### Algoritmo: Recomendación

```

Input: modelosKOS, paginasPortal, idiomaPortal
Output: rankingKOS
# Transformar sitio web en una colección de documentos
1: textosPortal = {}
2: for paginaHTML in paginasPortal do
3: textoPagina = transformacionATextoPlano(paginaHTML)
4: textosPortal = add(textosPortal, textoPagina)
5: end for
# Transformar modelos KOS en una colección de documentos
6: textosKOS = {}
7: for modelo in modelosKOS do
8: if modelo contiene etiquetas en idiomaPortal then
9: textoKOS = ""
10: for concepto in modelo do
11: etiquetas = {'prefLabel', 'altLabel', 'definition', 'scopeNote'}
12: for etiqueta in etiquetas do
13: textoKOS = append(textoKOS,
    extraer(modelo, concepto, etiqueta, idiomaPortal))
14: end for
15: textosKOS = add(textosKOS, <textoKOS, modelo>)
16: end for
17: N = length(textosKOS)
18: end if
19: end for
# Inicialización de puntuaciones de cada modelo KOS
20: rankingKOS = {}
21: for elemento in textosKOS do
22: modelo = elemento.getValor()
23: rankingKOS.put(modelo, 0)
24: end for
25: indiceModelosKOS = indexar(textosKOS)
# Votación de cada página a los modelos KOS
26: for textoPagina in textosPortal do
27: consulta = parsear(textoPagina)
28: resultados = buscarModeloVectorial(consulta, indiceModelosKOS)
29: puntuacion = N
30: for modelo in resultados do
31: rankingKOS.put(modelo, rankingKOS.get(modelo) + puntuacion)
32: puntuacion = puntuacion - 1
33: end for
34: end for
# Ordenar en sentido decreciente los modelos según puntuación
35: rankingKOS = sort_descending(rankingKOS)
36: return rankingKOS

```

Figura 5. Algoritmo de recomendación de modelos KOS

En este paso del procedimiento se pretende establecer un ranking de los modelos KOS más cercanos al contenido de las páginas web del portal que queremos anotar. Para ello se propone adaptar el modelo vectorial de recuperación de información que proporciona un ranking de resultados según la cercanía entre los vectores que representan a los documentos y el vector que representa a la consulta. Utilizando el coseno del ángulo formado entre el vector documento y el vector consulta se devuelve en primer lugar los vectores documento que forman un ángulo cercano a cero respecto al vector consulta. Adaptando el modelo vectorial a nuestro contexto, tal como se indica en la Figura 4 en nuestro caso cada modelo KOS es un documento y las páginas individuales actúan como consultas del sistema de recuperación.

La Figura 5 muestra el algoritmo seguido para establecer un ranking de los modelos KOS recomendados para el contenido del portal. En primer lugar, se procesa cada una de las páginas del portal para transformarlas a texto plano y convertir así el sitio web en una colección de documentos (líneas 1-5). A continuación, se hace un filtrado de los modelos KOS disponibles en el idioma del portal y aquellos que pasan el filtro se transforman también en un documento de texto (líneas 6-19). Para esta transformación a un documento de texto se extraen las etiquetas preferidas (*prefLabel*), las etiquetas alternativas (*altLabel*), las definiciones (*definition*) y las notas de alcance (*scopeNote*). Por último, se establece un sistema de votación de los modelos KOS (líneas 20-36). Para cada página del portal se calculan los modelos KOS más cercanos según el modelo vectorial indicado anteriormente y se otorga una puntuación a los modelos KOS. Si hay N modelos KOS para puntuar, el KOS más cercano a la página recibe N puntos, el siguiente KOS recibe N-1 puntos y así sucesivamente.

Finalmente, el algoritmo devuelve un listado ordenado de los modelos KOS según el sumatorio de las puntuaciones otorgadas por cada página. Respecto al procesamiento de texto utilizado en las líneas 25 y 27 para representar los modelos KOS y las páginas del portal como vectores de palabras, cabe destacar que se ha utilizado un analizador estándar para castellano proporcionado por el motor de indexación Lucene a través de la clase *SpanishAnalyzer*, la cual realiza una conversión a minúsculas de las palabras contenidas en el flujo de entrada de texto, separa los *tokens* con caracteres que no se corresponden con letras, elimina un conjunto básico de palabras vacías en castellano, y extrae las raíces de las palabras con un algoritmo de *stemming* (Savoy, 2002).

### 2.3. Anotación del contenido del portal.

Una vez que ya sabemos cuál es el modelo KOS más adecuado para anotar el contenido del portal, el último paso es procesar el texto de cada página e incluir los metadatos con los conceptos del modelo KOS que mejor se ajustan a ese texto.

---

#### Algoritmo: Anotacion

---

```

Input: modeloKOS, paginasPortal
Output: paginasAnotadas
1: paginasAnotadas = {}
2: grafoRDF = cargarGrafo(modeloKOS)
3: for pagina in paginasPortal do
4:   tiposEntidad = {'subject', 'location', 'organization', ...}
5:   anotaciones = {}
6:   for tipoEntidad in tiposEntidad do
7:     entidades = extraerEntidades(pagina, tipoEntidad)
8:     for entidad in entidades do
9:       etiquetas = {'prefLabel', 'altLabel', 'definition', 'scopeNote'}
10:      concepto = grafoRDF.query(entidad, tipoEntidad, etiquetas)
11:      end for
12:      anotaciones = add(anotaciones, concepto)
13:    end for
14:    paginasAnotadas = add(<pagina, anotaciones>)
15:  end for
16: return paginasAnotadas

```

---

Figura 6. Algoritmo de anotación

La Figura 6 indica el procedimiento seguido para realizar la anotación de cada página. En primer lugar, se carga el grafo RDF correspondiente al modelo KOS seleccionado (línea 2) para poder realizar consultas sobre él utilizando el lenguaje SPARQL. Posteriormente, para cada página, se extraen las expresiones correspondientes a entidades nombradas, o simplemente entidades, que hacen referencia a materias, lugares u organizaciones (línea 7) con una librería de procesamiento de lenguaje natural como Apache OpenNLP (Apache, 2017). A continuación, utilizando esas expresiones se buscan conceptos del modelo KOS que tengan algún emparejamiento textual en sus etiquetas *prefLabel*, *altLabel*, *definition* o *scopeNote* utilizando SPARQL en combinación con el motor de indexación Lucene. Además, se tiene en cuenta que el modelo KOS haya podido definir esquemas de conceptos separados para cada tipo de entidad (materia, lugar, etc.).

### 3. Experimentos iniciales

Tal como se comentó en la introducción, el método propuesto en este trabajo se ha aplicado al portal InfoGoya (Universidad de Zaragoza, 1996), un catálogo virtual de las obras del pintor Goya. Este portal consta de 2.398 páginas HTML: 1.576 páginas en castellano y 822 páginas con la traducción al inglés de los elementos principales del portal. Para el propósito de este

experimento nos hemos centrado exclusivamente en el uso de las páginas disponibles en castellano.

Para el primer paso de identificación y análisis de modelos KOS, se partió directamente de los resultados obtenidos por Díaz-Corona et al. (2019) donde se analizaron 5 repositorios de datos enlazados de relevancia en el dominio cultural, entre ellos Europeana (Comisión Europea, 2021) y la Digital Public Library of America (DPLA, 2021), y como resultado final se identificaron 16 modelos KOS con formatos compatibles con SKOS utilizados por un gran número de proveedores.

En la Tabla I se muestran los 8 modelos KOS disponibles en castellano con detalles sobre el número de conceptos en cada idioma (por ejemplo, columna *PL(es)* para castellano), relaciones BT-NT (*broader term – narrower term*), relaciones RT y la valoración global de su calidad. Como se puede observar, los modelos KOS se muestran ordenados de mayor a menor calidad. En general, la calidad global de la construcción de los modelos KOS es alta: 7 de los 8 modelos superan el 80% y el modelo mejor valorado es el tesoro GEMET, un tesoro propuesto en el dominio de medio-ambiente pero que también incluye términos generales en otros dominios. Aunque no es propiamente un tesoro cultural, este tesoro se utiliza en Europeana para interrelacionar recursos de clasificados con distintos modelos KOS.

Después de aplicar el sistema de votación descrito en la sección 2.2 (algoritmo de la Figura 5) a los modelos KOS candidatos, se obtuvo el ranking de puntuaciones que se muestra en la Figura 7. La puntuación de cada modelo KOS también se muestra en la columna *Puntuación* de la Tabla I.

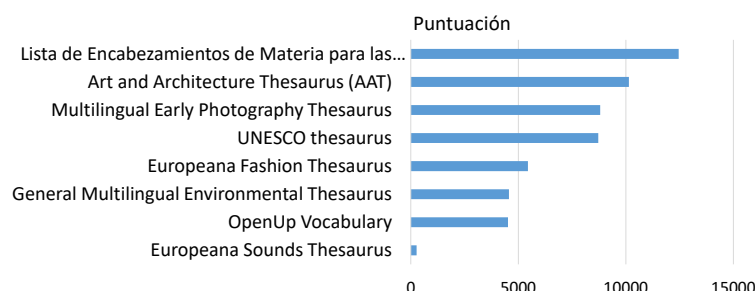


Figura 7. Ranking de modelos KOS

Como se puede observar, el modelo ganador es la Lista de Encabezamientos de Materias para las Bibliotecas Públicas (LEmb). Este modelo ha sido elaborado por la Subdirección General de Coordinación de Bibliotecas, y consta de 19.152 concep-

tos con etiquetas preferidas, etiquetas alternativas y notas de alcance en castellano. El ranking de modelos KOS difiere completamente de la ordenación por calidad mostrada en la Tabla I.

Nuestro sistema de recomendación ha conseguido colocar a los modelos sobre cultura y arte en las primeras posiciones y desplazar a los modelos del dominio de las ciencias naturales como GEMET y Open Up a las últimas posiciones.

Modelo KOS	#	PL(en)	PL(es)	PL(fr)	BT-NT	RT	Calidad	Puntuación
General Multilingual Environmental Thesaurus (Gemet)	5.244	5.235	5.236	5.235	6.555	2.086	94.59%	4.566
Lista de Encabezamientos de Materia para las Bibliotecas Públicas (LEMB)	19.152	0	19.142	0	7.227	24.875	88.48%	12.456
Europeana Fashion Thesaurus	1.089	1.088	799	1.088	1.147	0	87.34%	5.450
Europeana Sounds Thesaurus	15	15	15	15	9	0	87.00%	268
UNESCO Thesaurus	4.424	2.952	3.431	3.295	3.192	12.015	86.16%	8.718
Multilingual Early Photography Thesaurus	553	553	553	553	0	671	84.55%	8.803
Art and Architecture Thesaurus (AAT)	45.942	42.508	31.078	3.548	369.170	18.165	81.28%	10.142
OpenUp Vocabulary	733.097	102.077	19.640	1.195	0	0	61.56%	4.523

Tabla I. Modelos KOS candidatos para la anotación del portal

Por último, se ha seguido el procedimiento propuesto para la anotación (algoritmo incluido en la Figura 6, y explicado en la sección 2.3) con el objetivo de enriquecer las páginas en castellano del portal InfoGoya con conceptos del modelo LEMB. Para cada página se han extraído entidades nombradas de tipo materia y de tipo lugar dado que los conceptos de LEMB se distribuyen en dos esquemas separados: 'Encabezamientos de materia' y 'Encabezamientos de lugar'. De las 1.576 páginas en castellano, se han anotado de forma automática 1.075 páginas (68,21% del total en castellano) con una media de 2,34 conceptos por página. En la Tabla II se muestra el ejemplo de la anotación de una página que contiene el contenido de una carta enviada por el pintor en castellano antiguo donde se han reconocido dos entidades de lugar (*Madrid* y *San Luis*) y dos entidades referentes a una materia (*Madre* y *Jaqueca*) que se han enlazado con cuatro conceptos de LEMB. Se puede ver que la precisión no es perfecta ya que solo el 50% de las entidades se emparejan con conceptos correctos de LEMB. Sin embargo, una verificación manual exhaustiva de las 1.075 páginas donde se han extraído un total de 2.658 entidades mejora esta precisión, obteniéndose que el 79,21% las entidades extraídas se emparejaron correctamente con un concepto de LEMB. De todas formas, hay que tener en cuenta que esta precisión es mejorable ya que de momento no se ha incluido ningún mecanismo desambiguación. Simplemente, se ha escogido

el primer concepto de LEMB que consigue la mejor valoración en la correspondencia textual de Lucene para las cuatro etiquetas mencionadas en la sección 2.2.

Tipo Entidad	Texto	Concepto LEMB
Lugar	Madrid	Madrid
	San Luis	San Luis Potosí (Estado)
Materia	Madre	Madres trabajadoras
	Jaqueca	Cefalalgia

Tabla II. Ejemplo de anotación de una página del portal (<http://goya.unizar.es/Repositorio/Diplomatario/108.html>)

#### 4. Conclusiones

Este trabajo ha presentado un método sencillo para recomendar un modelo KOS entre un conjunto de modelos KOS candidatos, y cómo utilizar el modelo recomendado para la anotación de las páginas de un portal. Para la recomendación del modelo KOS se ha reaprovechado el potencial de modelos clásicos de recuperación de información, véase el modelo vectorial, con el objetivo de ver la correspondencia entre los contenidos que se quieren anotar y el contenido textual de los modelos KOS. Para la parte de anotación de las páginas se ha reutilizado el software disponible de reconocimiento de entidades nombradas y se ha hecho uso del lenguaje SPARQL combinado

con motores de indexación textual como Lucene para encontrar correspondencias entre entidades nombradas y etiquetas textuales asociadas a los conceptos del modelo KOS elegido.

Para verificar la viabilidad de este modelo se ha implementado un prototipo inicial y se ha aplicado sobre el portal InfoGoya. Aunque algunos conceptos extraídos automáticamente del modelo LEMb pueden no ser demasiado precisos, proporcionan una buena base para que los administradores del portal puedan anotar semánticamente el mismo.

Como trabajo futuro proponemos mejorar nuestro proceso de identificación y análisis de modelos KOS con la integración de bases de datos de modelos KOS de relevancia como BARTOC (Verbundzentrale des GBV, 2021) para aquellos casos en los que no se cuente con un trabajo previo de análisis de modelos KOS utilizados en un dominio particular. Asimismo, estudiaremos otras técnicas de clasificación para asignar el KOS más cercano al contenido de un portal. También se trabajará en la integración de mecanismos de desambiguación que mejoren el emparejamiento entre entidades reconocidas en el texto y los conceptos del modelo KOS. Por último, se plantea estudiar distintas alternativas para integrar las anotaciones como micro-metadatos RDF embebidos en las páginas HTML del portal semantizado.

## Agradecimientos

Este trabajo ha sido financiado parcialmente por el Gobierno de Aragón (proyecto T29\_20R).

## Referencias

- Albertoni, R.; De Martino, M.; Quarati, A. (2016). Integrated quality assessment of linked thesauri for the environment. // International Conference on Electronic Government and the Information Systems Perspective. 221–235.
- Apache (2017). The Apache Software Foundation. Apache OpenNLP web site. <https://opennlp.apache.org/> (2021-03-24)
- Comisión Europea (2021). Portal de Europeana. <https://www.europeana.eu/> (2021-03-24)
- da Silva Lemos, D.L.; Souza, R.R. (2020). Knowledge Organization Systems for the Representation of Multimedia Resources on the Web: A Comparative Analysis. // Knowledge Organization. 47:4, 300-319.

- Daiber, J.; Jakob, M.; Hokamp, C.; Mendes, P. N. (2013). Improving efficiency and accuracy in multilingual entity extraction. // Proceedings of the 9th International Conference on Semantic Systems. 121-124.
- Díaz-Corona, D.; Lacasta, J.; Latre, M.Á.; Zarazaga-Soria, F.J.; Nogueras-Iso J. (2019). Profiling of knowledge organisation systems for the annotation of Linked Data cultural resources. // Information Systems. 84, 17-28.
- DPLA (2021). Portal de la Digital Public Library of America. <https://dp.la/> (2021-03-24)
- Khalili, A., Auer, S. (2015). WYSIWYM–Integrated visualization, exploration and authoring of semantically enriched un-structured content. // Semantic Web. 6:3, 259-275.
- Lacasta, J.; Falquet, G.; Zarazaga-Soria, F.J.; Nogueras-Iso, J. (2016). An automatic method for reporting the quality of thesauri. // Data & Knowledge Engineering. 104, 1-14.
- Lacasta, J.; Nogueras-Iso, J.; Zarazaga-Soria, F. J. (2010). Terminological Ontologies: Design, Management and Practical Applications. Springer.
- Miles, A.; Brickley, D. (2005) SKOS Core Guide. <http://www.w3.org/TR/swbp-skos-core-guide> (2021-03-24).
- Rodriguez Rocha, O.; Vagliano, I.; Figueroa, C.; Cairo, F.; Futia, G.; Licciardi, C. A.; Marengo, M.; Morando, F. (2015). Semantic annotation and classification in practice. // IT Professional. 17:2, 33-39.
- Savoy J. (2002) Report on CLEF-2001 Experiments: Effective Combined Query-Translation Approach. // Peters, C.; Braschler, M.; Gonzalo, J.; Kluck, M. (eds). Evaluation of Cross-Language Information Retrieval Systems. CLEF 2001. Lecture Notes in Computer Science. 2406, 27-43.
- Suominen, O.; Mader, C. (2014). Assessing and improving the quality of SKOS vocabularies. // Journal of Data Semantics. 3:1, 47–73.
- Tchechmedjiev, A.; Abdaoui, A.; Emonet, V.; Melzi, S., Jonnagaddala, J.; Jonquet, C. (2018). Enhanced functionalities for annotating and indexing clinical text with the NCBO Annotator+. // Bioinformatics. 34:11, 1962-1965.
- Thessen, A.E.; Parr, C.S. (2014). Knowledge extraction and semantic annotation of text from the encyclopedia of life. // PloS one. 9:3, e89550.
- Universidad de Zaragoza (1996). Exposición Virtual InfoGoya '96. <http://goya.unizar.es/> (2021-03-24)
- Verbundzentrale des GBV (2021). Basic Register of Thesauri, Ontologies & Classifications (BARTOC). <https://bartoc.org/> (2021-06-08).
- Whaley, P.; Edwards, S.W.; Kraft, A.; Nyhan, K.; Shapiro, A.; Watford, S.; Wattam S.; Wolffe, T.; Angrish, M. (2020). Knowledge Organization Systems for Systematic Chemical Assessments. // Environmental Health Perspectives. 128:12, 125001.

Enviado: 2021-03-25. Segunda versión: 2021-06-09.  
Aceptado: 2021-06-17.



---

# Contar historias con los datos: Aragón Open Data Focus, una experiencia innovadora de reutilización de los datos del sector público

*Data storytelling: Aragón Open Data Focus, an innovative experience of reusing public sector data*

---

Lorea AKERRETA ESCRIBANO (1), Julián MOYANO COLLADO (2)

(1) Deloitte, Plaza del Castillo, 21, 2º, 31008 Pamplona, Navarra (España), lakerreta@deloitte.es.

(2) Gobierno de Aragón, Edificio Pignatelli, Paseo María de Agustín, 36, 50071 Zaragoza (España), jmoyano@aragon.es

## Resumen

En los últimos años, las Administraciones Públicas han realizado, en un ejercicio de transparencia, un importante esfuerzo por abrir sus datos para que, a través de la reutilización de esta información, pueda generarse un impacto y valor real sobre el territorio. Sin embargo, en muchas ocasiones su complejidad hace que la ciudadanía no perciba los beneficios de su uso y la utilice. En este escenario, los avances tecnológicos y una nueva cultura de consumo de la información derivada de la democratización en su acceso han sentado las bases del desarrollo del periodismo de datos. Aragón Open Data, la apuesta de Gobierno de Aragón por los datos abiertos, ha desarrollado durante 2020 un nuevo servicio que busca unir estos dos factores con el objetivo de incentivar la reutilización de los datos del sector público, hacer pedagogía sobre sus posibilidades y empoderar a los distintos perfiles de usuarios en el uso de los datos públicos. El resultado del proyecto, enmarcado en su estrategia 2019-2022 y cofinanciado por el Fondo Europeo de Desarrollo Regional (FEDER) de la Unión Europea, es Aragón Open Data Focus, un servicio gratuito que permite a los usuarios del portal desarrollar y compartir sus propias historias a partir de los datos abiertos de Aragón.

**Palabras clave:** Datos abiertos. Data storytelling. Transparencia. Reutilización de la información. Innovación pública. Aragón.

## 1. Introducción

La Open Knowledge Foundation (2012), fundación sin ánimo de lucro creada en 2004 para apoyar la difusión de conocimiento abierto en sentido amplio, define los datos abiertos como

[...] datos que pueden ser utilizados, reutilizados y redistribuidos libremente por cualquier persona, y que se encuentran sujetos, cuando más, al requerimiento de atribución y de compartirse de la misma manera en que aparecen.

En el contexto de la tendencia observada en los últimos años hacia modelos de gobierno abierto, más transparentes, participativos y orientados a la rendición de cuentas a la ciudadanía, los datos

## Abstract

In recent years, Public Administrations have made, in an exercise of transparency, an important effort to open their data with the aim that, through the reuse of this information, a real impact and value can be generated on the territory. However, on many occasions the complexity of the matter means that citizens are not able to perceive the benefits that can be derived from its use, and they distance themselves from the use of open data. In this same scenario, technological advances and a new culture of information consumption derived from the democratisation of its access have laid the foundations for the development of data journalism. Aragón Open Data, the Government of Aragón's commitment to open data, has developed a new service during 2020 that seeks to unite these two factors with the aim of encouraging the reuse of public sector data, educating about its possibilities and empowering the different user profiles in the use of public data. The result of the project, framed in its 2019-2022 strategy and co-financed by the European Regional Development Fund (ERDF) of the European Union, is Aragón Open Data Focus, a free service that allows users of the portal to develop and share their own stories based on Aragón's open data.

**Keywords:** Open data. Data storytelling. Transparency. Information reuse. Public innovation. Aragón.

abiertos han ido escalando en términos de importancia como uno de los elementos nucleares de este tipo de políticas. Los distintos gobiernos han invertido una importante cantidad de recursos en abrir la información de la que disponen para que cualquier persona interesada pueda acceder a ella y reutilizarla con la única obligación de reconocer su autoría y citar al organismo publicador como fuente de dichos datos. El caso de España es especialmente notable, situándose año tras año en las primeras posiciones de la Unión Europea en los estudios de madurez de apertura de datos del EU Open Data Portal en los cuatro indicadores observados: política, portal web, impacto y calidad.

Sin embargo, este esfuerzo por parte de las Administraciones Públicas se ha visto, en la práctica, poco correspondido por los usuarios. Con la excepción de colectivos específicos de usuarios, como programadores y desarrolladores, en general se desconoce la existencia de estas iniciativas o no se hace uso de ellas por desinterés en la información publicada, percepción de complejidad o, simplemente, porque no se aprecia el valor de la reutilización de estos datos.

En este sentido, algunas Administraciones Públicas se encuentran inmersas en la búsqueda de soluciones o alternativas que les permitan afrontar el reto de acercar los datos abiertos al común de la ciudadanía y facilitar su reutilización. Este es el caso del Gobierno de Aragón, que en su Estrategia Aragón Open Data 2019-2022 (Gobierno de Aragón, 2019) establece entre sus siete ejes de actuación el potenciar la implicación de los usuarios del portal e incluir nuevos servicios de valor añadido.

Bajo esta premisa, la Dirección General de Administración Electrónica y Sociedad de la Información de Gobierno de Aragón, responsable de la administración del portal autonómico de datos abiertos, llevó a cabo un proceso de reflexión interno en el que, mediante la observación de experiencias de éxito a nivel internacional en materia de reutilización de la información pública, identificó como buena práctica capaz de generar un retorno de valor real la puesta a disposición de un servicio que permitiera que usuarios sin conocimientos técnicos pudieran manipular los datos abiertos disponibles para generar visualizaciones que pudieran emplearse, a su vez, para construir historias en las que se interpretaran estos. Los beneficios de un servicio de este tipo van desde el empoderamiento de la ciudadanía al facilitar su acceso a la información pública hasta reforzar la transparencia de la actuación pública, pasando por el establecimiento de un marco de diálogo directo con distintos tipos de usuarios que permita mejorar, además, la calidad del servicio prestado.

Así, en el año 2020, el Gobierno de Aragón se puso manos a la obra y lanzó el proyecto Aragón Open Data Focus, nombre elegido para el nuevo servicio del portal de datos abiertos de Aragón. El trabajo desarrollado partió de un análisis de los datos disponibles y la caracterización de los potenciales colectivos de usuarios que podrían hacer uso de estos datos en función de su conocimiento de los datos abiertos, su actitud hacia los mismos o los temas que pudieran resultarles de interés. Tras este análisis, se abrió un proceso de diálogo, condicionado por la pandemia de la COVID-19, con los distintos colectivos definidos.

The image shows a screenshot of the 'ARAGÓN OPEN DATA' portal's 'CORONAVIRUS EN ARAGÓN' dashboard. The dashboard is titled 'ARAGÓN OPEN DATA' and 'OPEN DATA FOCUS'. The main content area has a blue background and features a line chart titled 'Radiografía de los casos de Coronavirus confirmados por provincia de la Comunidad Autónoma de Aragón'. Below this is a video player titled 'El proceso de cuarentena'. Further down is a multi-line chart titled 'Evolución de la ocupación de camas COVID-19 por hospital'. To the right of this chart is a text block titled 'El rol del sistema aragonés de salud en la lucha contra el Coronavirus'. Below that is another text block titled 'Fondo Covid Aragón'. At the bottom of the main content area is a line chart titled 'Aulas afectadas por coronavirus en Aragón, por provincia'. The footer contains 'AVISO LEGAL', 'CONTENIDO DESTACADO', 'POWERED BY' logos, and the 'GOBIERNO DE ARAGÓN' logo.

Figura 1. Historia “Coronavirus en Aragón”



En este proceso participaron casi 300 personas a través de distintos eventos web en los que se daba a conocer el proyecto de Aragón, se presentaban distintos casos de interés adaptados a la caracterización de cada colectivo, se abría un debate con los participantes sobre los datos que les resultaban de mayor interés y se identificaban los aspectos de mejora que les facilitarían reutilizar la información disponible.

El resultado de este trabajo ha sido la creación de un servicio web que permite a los usuarios del portal de Aragón convertirse en “periodistas de datos”, utilizando los más de 5.000 recursos disponibles en Aragón Open Data para la construcción de historias de su interés y consultar las generadas por otros usuarios. Así, a través del servicio, los usuarios pueden construir, de una forma sencilla, visualizaciones con los datos del catálogo, acompañarlas de recursos adicionales (como vídeos, imágenes o presentaciones) y redactar sus propios artículos con base en la información disponible. La Figura 1 muestra una de las historias creadas en el nuevo servicio que explica cómo ha impactado la pandemia de coronavirus en Aragón, a través de la presentación gráfica de los casos registrados por provincia, la evolución de la ocupación de camas COVID-19 por hospital y las aulas afectadas por provincia, además de un vídeo que explica el proceso de cuarentena para evitar la expansión del virus.

El presente artículo pretende presentar el proyecto desarrollado desde Gobierno de Aragón, profundizando en el potencial de los datos abiertos para crear valor, presentando las conclusiones obtenidas del trabajo desarrollado con los distintos colectivos y explicando el funcionamiento del nuevo servicio, poniendo en valor esta experiencia pionera en España y abriendo la puerta a nuevos modelos de innovación en el acercamiento de los datos abiertos a la ciudadanía. Para ello, se llevará a cabo una breve presentación del contexto actual de los datos abiertos en el sector público y su contribución al desarrollo económico, la mejora de los servicios públicos y el empoderamiento de la ciudadanía por parte de los gobiernos abiertos, seguida de una explicación del proyecto Aragón Open Data Focus y los resultados alcanzados, para finalizar con las principales conclusiones obtenidas de los trabajos desplegados.

## 2. Contexto

### 2.1. Antecedentes: la transparencia de las Administraciones públicas

La pérdida de confianza de la ciudadanía en las instituciones públicas y sus gestores ha supuesto

un catalizador para la creación de nuevos mecanismos de relación que permitan a la ciudadanía conocer cómo se toman las decisiones, cómo se manejan y gestionan los fondos públicos o cuáles son los criterios que rigen la actuación de los gobiernos (Cerrillo i Martínez, 2011).

En este contexto, la transparencia, frente a la idea de secretismo y de reserva de la información, se erige como una herramienta clave para asegurar la higiene democrática, así como la participación e involucración de la ciudadanía en la gestión de lo público, constituyéndose como una de las reglas de la legitimación real y social del comportamiento de los gobiernos, instituciones públicas y clase política (Cordero, 2007).

La normativa, tanto de carácter internacional, como a nivel estatal, avala y aboga por una transparencia real de las Administraciones públicas. Así, la Carta de Derechos Fundamentales de la Unión Europea, en sus artículos 41 y 42, reconoce el derecho de los ciudadanos europeos a una buena administración y a acceder libremente a los documentos contenidos en las instituciones europeas (Palomares, 2017), sentando las bases para el desarrollo normativo de este principio.

Asimismo, el Convenio 205 del Consejo de Europa sobre el Acceso a los Documentos Públicos, de 18 de junio de 2009, reconoce la importancia de la transparencia de autoridades públicas en una sociedad democrática y plural (Montero, 2020), considerando que el ejercicio del derecho de acceso a los documentos públicos proporciona una fuente de información para los distintos tipos de agentes que viven y operan en los territorios; ayuda al público a formarse una opinión sobre el estado de la sociedad y sobre la labor de la Administración; y fomenta la integridad, la eficacia, la eficiencia y la responsabilidad de autoridades públicas, ayudando así a que se afirme su legitimidad. Considera, por tanto, que todos los documentos generados en el ejercicio de las competencias de la Administración son en principio públicos y solamente pueden ser retenidos para proteger otros derechos e intereses legítimos.

En el ámbito estatal y con carácter básico, se aprobó la Ley 19/2013, de 9 de diciembre, de transparencia, de acceso a la información pública y buen gobierno. Su aplicación y contenido se vertebra en tres vertientes: la transparencia en la actividad pública a través de la publicidad activa; el reconocimiento y la garantía del derecho de acceso a la información pública; y el buen gobierno, que se concreta en obligaciones que vinculan a los responsables públicos y en las consecuencias que derivan de su vulneración (Cotino, 2015).

Así, si bien es cierto que existe una norma jurídica que obliga a la transparencia de la información por parte de las Administraciones públicas, las mismas están aprovechando las nuevas tecnologías para incrementar exponencialmente la difusión pública de toda la información de interés general que poseen las entidades que se financian con los impuestos de los ciudadanos, yendo más allá, buscando fórmulas que les permitan configurar gobiernos más abiertos y accesibles (Blanes, 2013).

## 2.2. La reutilización de la información pública como materialización del concepto de Gobierno Abierto

Consecuencia de la evolución de las relaciones con la ciudadanía, emerge el concepto de Gobierno Abierto, que adquiere carta de naturaleza institucional en 2009 con la aprobación de una directiva de gobierno abierto, la *Open Government Directive* (Obama, 2009), por la Administración Obama en Estados Unidos, como un nuevo enfoque aglutinador de esfuerzos por mejorar las capacidades del gobierno y mejorar las administraciones públicas (Conejero, 2013), estableciéndose como una tipología de gobernanza pública basada en los principios de transparencia, colaboración y participación. Se trata de un modelo de gobernanza que apuesta por el diálogo y el equilibrio de poder entre gobierno y gobernados, al mismo tiempo que aboga por la corresponsabilidad ciudadana en la gestión pública (Clabo y Ramos-Vielba, 2015).

En este escenario aparece la apuesta de los gobiernos por los datos abiertos como un nuevo camino a explorar en las relaciones con la ciudadanía. Se distinguen de la transparencia en que mientras que esta consiste en la puesta a disposición de información relevante de una manera comprensible para asegurar que cumple su propósito divulgativo, los datos abiertos presentan la información de la manera menos tratada, en distintos formatos y con el mayor grado de desagregación disponible, de manera que puedan ser utilizados o reutilizados con otros propósitos que el meramente informativo, siempre con el único requisito de atribución de su fuente (Aleixandre-Benavent, Ferrer y Peset, 2019), contribuyendo a la satisfacción de distintas necesidades particulares de los usuarios.

En este contexto, el sector público produce una gran variedad de información que es potencialmente reutilizable, como puede ser la información social, económica, geográfica, estadística, etc. Esta información es muy atractiva para su reutilización ya que es completa, fiable y de calidad, estimando el *European Data Portal*, en su

estudio *The Economic Impact of Open Data: Opportunities for Value Creation in Europe* (European Data Portal, 2020), que el tamaño de mercado para los datos abiertos tiene un valor de 184.000 millones de euros y que, para 2025, crecerá hasta una horquilla de entre 199.000 y 334.000 millones de euros. Así, a través de iniciativas de Open Data, se quiere lograr que la reutilización y redistribución de la información pública repercuta en beneficio tanto de las administraciones públicas (Abella, 2019), que sacan a la luz sus datos, como de los ciudadanos y empresas, que ven cubiertas algunas de sus necesidades reales.

Si bien existen numerosos ejemplos de las formas en que los datos abiertos ya están creando valor social y económico, aún se desconoce la amplitud completa de sus posibilidades a futuro. Las nuevas combinaciones de datos pueden crear nuevos conocimientos y perspectivas, que pueden conducir a campos de aplicación completamente nuevos. El pasado nos enseña, una vez más, que la combinación de datos nuevos da lugar a un conocimiento mayor: por ejemplo, cuando el Dr. John Snow descubrió la relación entre la contaminación del agua potable y el cólera en Londres en el siglo XIX al combinar datos sobre las muertes por cólera con la ubicación de los pozos de agua. Esto condujo a la construcción de los sistemas de alcantarillado de Londres y mejoró enormemente la salud general de la población. Desarrollos como este continuarán sucediendo a medida que fluyan conocimientos inesperados de la combinación de diferentes conjuntos de datos abiertos.

Este potencial sin explotar se puede desatar si los datos públicos del gobierno se convierten en datos abiertos. Sin embargo, esto solo sucederá si los datos son real y efectivamente abiertos, es decir, si no existen restricciones (legales, financieras o técnico-tecnológicas) para su reutilización por parte de otros usuarios: la existencia de cualquier obstáculo excluirá a las personas de reutilizar los datos públicos y dificultará la búsqueda de formas valiosas de hacerlo.

## 2.3. La contribución al desarrollo económico y competitivo de los territorios, la mejora en la prestación de los servicios públicos y el empoderamiento de la ciudadanía por parte de los gobiernos abiertos

Los datos abiertos contribuyen directamente al crecimiento de las economías y a la transformación de la sociedad, y buena prueba de ello es la apuesta e implicación de gobiernos de todo el mundo por la promoción de portales que hagan accesibles los datos públicos más allá de servir

como tarjeta de presentación para exhibir la transparencia pública, de bajo coste y relativamente fácil de construir (Ramos, et al., 2012).

En el contexto de la innovación, la puesta a disposición de datos abiertos por parte de la Administración pública estimula que se descubran nuevas oportunidades de negocio para empresas y se eliminen barreras al acceso, uso e intercambio de datos, de manera que finalmente se contribuya al crecimiento económico (Attard, et al., 2015). Por ejemplo, en Islandia se publican con carácter anual datos correspondientes a la calidad de los corderos de las distintas granjas de cría existentes en el país. De esta manera, desde las instituciones públicas, se facilita a granjeros, agricultores u otros agentes de la cadena de valor, información que puede ser de su interés.

Asimismo, el hecho de poner en disposición de empresas los datos fomenta un mayor conocimiento, contribuyendo en última instancia a que las decisiones empresariales se tomen con mayor base, resultando estas más adecuadas (Kucera y Chlapek, 2014).

A nivel gubernamental, los datos abiertos contribuyen a que los gobiernos puedan equilibrar mejor la prestación de servicios, abordando problemas como el tráfico, la seguridad, la movilidad en las ciudades u otras cuestiones, como la salud o la igualdad, transformando el modelo de relación con el ciudadano, creando sociedades más avanzadas y participativas, y fomentando la colaboración público-privada. Por ejemplo, en Reino Unido no existe una obligación legal para las autoridades locales de ofrecer baños públicos o información sobre sus equipamientos, localizaciones u horarios, lo que puede impactar sobre la igualdad o diversidad de la sociedad: un estudio de 2003 identificó que muchas mujeres planificaban sus desplazamientos diarios en función de los aseos públicos disponibles y conocidos (Greed, 2003). A esta circunstancia se sumó la aprobación en 2004 de la *Disability Discrimination Act* (DDA), que reconoció la incontinencia como una discapacidad. La solución a esta situación se articuló en 2011 a través del *Great British Public Toilet Map*, una iniciativa público-privada basada en datos abiertos que ofrece información sobre los aseos públicos existentes, mejorando la calidad de vida de la ciudadanía, especialmente entre la población de mayor edad (Richard y Knight, 2012).

Por otra parte, la existencia de datos abiertos permite a los gobiernos tener un mayor conocimiento de la estructura de coste de gobiernos de características similares, de empresas, etc., con-

tribuyendo a una mejor gestión de los fondos públicos. Al respecto, el estudio del *European Data Portal The Economic Impact of Open Data: Opportunities for value creation in Europe* (European Data Portal, 2020a), cifra el ahorro en costes para la Administración Pública en 2020 en 1.700 millones de euros.

Finalmente, los datos abiertos también permiten el desarrollo de nuevos servicios innovadores. Un ejemplo ilustrativo es la experiencia Desafío modal, en México, en la que se simulan situaciones de traslado de un origen a un destino con diversas formas de locomoción y diferentes rutas a través de la reutilización de datos abiertos. Los resultados de esta simulación son, cuanto menos, curiosos: en hora punta, bicicletas, patines y motocicletas aparecen como los medios de transporte más eficientes, acumulando la bicicleta una diferencia de más de 24 minutos con el automóvil en Ciudad de México, o de 15 en el caso de Oaxaca (Sève y Redondo, 2019).

En el escenario descrito, el desarrollo de un modelo de compartición de datos y co-creación de servicios entre los sectores público y privado se constituye como una pieza clave de futuro (Cabezali, 2020). Así lo ha entendido la UE, que en su informe *Analytical Report 12: Business-to-Government Data Sharing* (European Data Portal, 2020b) concluye que el suministro de datos desde entidades privadas a públicas (B2G) para su reutilización fomentará el crecimiento, el empleo y la competitividad europea ahorrando costes, tiempo y vidas, estableciendo en consecuencia un plan de 6 pasos para avanzar en este camino hacia la “quinta libertad”: el movimiento libre de datos en el Mercado Digital Europeo.

Sin embargo, el despliegue de iniciativas de datos abiertos no está exenta de dificultades: así, actualmente, dos de los principales retos que enfrenta son, por un lado, la necesidad de desplegar una labor pedagógica con la ciudadanía y el tejido empresarial que les permita comprender el potencial de su uso; y, por otro, facilitar esa reutilización mediante la capacitación de los distintos grupos de agentes que viven y operan en los territorios (Álvarez et al., 2014).

### 3. Objetivos del proyecto

#### 3.1. La visión de los datos abiertos del Gobierno de Aragón: retos y respuesta

Aragón Open Data (1) es el medio a través del que se materializa el compromiso del Gobierno de Aragón de publicar sus datos de forma abierta. Este compromiso se origina en el ejercicio 2012 y tiene como objetivo fundamental la

creación de riqueza económica a través de la reutilización de la información y el fomento de la transparencia, lo cual dota de un carácter particular al proyecto al constituirlo como un planteamiento más evolucionado del proceso de apertura de la información pública que busca, ya desde sus inicios, el fomento de la utilización de la misma para la creación de un valor tangible.

A lo largo de los años se han sucedido avances notables en el proyecto, administrado por la Dirección General de Administración Electrónica y Sociedad de la Información de Gobierno de Aragón, que han permitido situar al portal de datos abiertos de Aragón a la cabeza de las propuestas autonómicas en España, tal y como se muestra en la Tabla I.

Comunidad Autónoma	Datasets	API	Catál. RDF
Andalucía	555	No	No
Aragón	2.823	Sí	Sí
Asturias	24	No	No
Baleares	182	Sí	Sí
Canarias	186	Sí	No
Cantabria	356	Sí	No
Castilla-La Mancha	266	No	No
Castilla y León	487	No	No
Cataluña	577	Sí	No
Comunidad Valenciana	492	Sí	No
Extremadura	25	No	No
Galicia	366	No	Sí
La Rioja	318	Sí	Sí
Madrid	432	Sí	No
Murcia	765	Sí	No
Navarra	1.380	Sí	No
País Vasco	4.966	Sí	No

Tabla I. Comparativa entre los portales de datos autonómicos, 2020

Así, en 2020 era la segunda Comunidad en términos de conjuntos de datos abiertos publicados, destacando su banco de datos (2) como uno de sus activos más importantes. Para facilitar la búsqueda de información en un repositorio tan complejo y heterogéneo, cuenta con un avanzado sistema de búsqueda que permite tanto la búsqueda libre como el empleo de otros criterios, ya sean el número de descargas del recurso, la fecha de

última actualización, temas, tipos de información, organización publicadora o etiquetas, entre otras.

La construcción de este banco de datos se ha materializado gracias a la implicación y compromiso de un amplio abanico de actores del sector público aragonés, que colaboran con el proyecto preparando y cediendo sus datos, desde 29 Direcciones Generales del gobierno autonómico (Administración Electrónica y Sociedad de la Información, Administración Local, Asistencia Sanitaria, Patrimonio y Organización, Cultura, Medio Natural y Gestión Forestal, etcétera) hasta la Universidad de Zaragoza, pasando por otras instituciones como los Institutos Aragoneses de Estadística, Empleo o del Agua, la Junta de Reclamaciones Económico-Administrativas de la Comunidad Autónoma de Aragón o Intervención General, entre otros. La variedad de publicadores enriquece las temáticas de información disponibles en el banco de datos, que pasan por cuestiones como ciencia y tecnología, comercio, cultura y ocio, demografía, deporte, economía, educación, empleo, energía, hacienda, industria, legislación y justicia, medio ambiente, medio rural, salud, sector público, seguridad, sociedad y bienestar, transporte, turismo o urbanismo e infraestructuras.

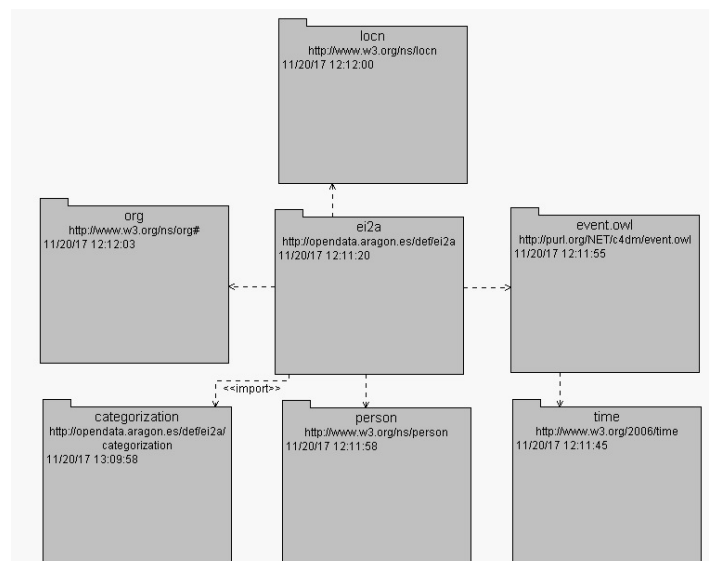


Figura 2. Ontologías y vocabularios usados para el desarrollo de la propuesta ontológica EI2A

Adicionalmente, Aragón Open Data ha desarrollado su propia ontología (EI2A, Estructura de Información Interoperable de Aragón), resumida en la Figura 2, que permite generar un lenguaje común entre las bases de datos y aplicaciones del Gobierno de Aragón, ha puesto a disposición de los usuarios del portal tres APIs adicionales a la nativa de CKAN (la aplicación web de código abierto que soporta el proyecto) para acceder a

los datos de Aragón Open Data y ha incorporado un punto SPARQL de acceso a la información semántica, con su correspondiente información de consulta, facilitando el uso de los datos por parte de los usuarios técnicamente más avanzados.

En este contexto de desarrollo del proyecto, el Gobierno de Aragón aprobó, en el ejercicio 2019, la Estrategia Aragón Open Data 2019-2022 con la visión de que el portal de datos abiertos del Gobierno de Aragón se constituya en el portal de referencia en el ámbito de la gestión del dato abierto y proporcione un tratamiento de los sistemas de información que no sólo genere riqueza, sino que desarrolle y consolide, entre otros, el sector infomediario. Dicha estrategia analizaba, además, el mapa actual de agentes reutilizadores de datos e interesados para obtener una visión integral de las distintas posibilidades de desarrollo. Este mapa conceptual puede consultarse en la Figura 3 de este artículo.



Figura 3. Mapa de agentes de Aragón Open Data

Con el objetivo de avanzar en su estrategia de datos abiertos, el Gobierno de Aragón ha considerado clave la participación para el buen funcionamiento de la iniciativa, y se han abierto canales donde los ciudadanos y empresas han podido trasladar sus necesidades en materia de datos. Con el fin de estimular las políticas de reutilización de la información del sector público, el Gobierno de Aragón trata de fomentar, mediante el impulso de iniciativas relevantes, el uso, reproducción, modificación o distribución de los conjuntos de datos.

Entre estas iniciativas destaca el proyecto Aragón Open Data Focus, lanzado en el año 2020, como una apuesta innovadora para ir un paso más allá en el desarrollo de los datos abiertos en el territorio. El objetivo principal del proyecto es construir un servicio que permita a cualquier persona crear narrativas digitales a partir de los datos abiertos disponibles en el catálogo de Aragón Open Data, poniendo a su disposición una herramienta que permita la reutilización de la informa-

ción de forma efectiva y la generación colaborativa de conocimiento sobre la realidad económica, social, demográfica y cultural del territorio de una forma visual y atractiva.

De esta manera, Aragón Open Data Focus se presenta como una respuesta del Gobierno de Aragón al reto de acercar la utilidad de los datos abiertos a la ciudadanía, tejido empresarial y resto de agentes del territorio. Su implementación y desarrollo ha sido cofinanciada por el Fondo Europeo de Desarrollo Regional (FEDER) de la Unión Europea, cuyo lema es “Construyendo Europa desde Aragón”.

### 3.2. Contar historias con los datos como valor diferencial de la apuesta de Aragón Open Data

Tal y como se ha adelantado, el proyecto Aragón Open Data Focus tiene su origen en la necesidad de mejorar el impacto del portal de datos abiertos Aragón Open Data. Para dar respuesta a este reto, se observaron distintas experiencias lanzadas en los últimos años en la esfera internacional en materia de reutilización de datos, identificándose como tendencia común el uso de visualizaciones para mejorar el entendimiento de la información que publican los poderes públicos. Entre estas experiencias analizadas destacan casos como Data USA (3), DataChile (4), Data Africa (5), el Departamento de Agricultura de EEUU (6) o el portal de datos abiertos de Singapur (7).

Los avances experimentados en los últimos años, tanto en accesibilidad y calidad de los datos como en el desarrollo de soluciones de tratamiento y visualización de estos, han provocado el surgimiento y auge del concepto *data storytelling* o “contar historias con los datos” como solución para la presentación, enriquecimiento y comprensión de la información en ámbitos tan diversos como la mejora de los procesos de toma de decisiones en el mundo empresarial (Suprata, 2019), la construcción de relatos publicitarios de marca (Caro-Castaño y Selva-Ruiz, 2019), la comunicación de análisis y resultados (Dykes, 2015) o el avance ético del periodismo (Lee et al., 2015).

La visualización de datos no es un elemento extraño a la estrategia de Aragón Open Data para avanzar en su labor sensibilización sobre el potencial de los datos abiertos: ya en 2018 desarrolló un servicio, Open Visual Data, que permitía a los usuarios del portal generar y utilizar sencillos gráficos con los datos disponibles en la web. Sin embargo, desde Gobierno de Aragón se ha llevado a cabo un esfuerzo adicional por configurar una propuesta de servicio un paso más allá, desarrollando para ello una herramienta de uso intuitivo que empodere a los usuarios menos

avanzados del portal en el uso de los datos públicos para construir sus propias narrativas digitales, apostando por un uso innovador del open data más vinculado a las narrativas visuales (Segel y Heer, 2010) y el periodismo de datos (Flores y Salinas, 2013).

#### 4. Metodología del proyecto Aragón Open Data Focus

El proyecto partió de una primera fase de análisis de los datos y servicios disponibles en el portal Aragón Open Data con el objetivo de conocer cuáles podrían ser los principales agentes interesados en cada uno de ellos, con el objetivo de mejorar la comunicación de su disponibilidad y facilitar su uso por parte de estos colectivos. Este diagnóstico de situación se estableció así como punto de partida que permitiera a futuro determinar la manera más efectiva y eficiente de aumentar el impacto del portal entre las distintas tipologías de agentes.

Así, en primer lugar, se definieron los distintos colectivos de potenciales usuarios del portal de Open Data de Aragón. Para ello, se examinaron distintas aproximaciones realizadas en otras iniciativas de datos abiertos en cuanto a la caracterización de los colectivos destinatarios de la información pública, identificándose como la práctica más interesante el trabajo desarrollado por la ciudad de Nueva York, *Meet the Users of Open Data* (NYC Open Data, 2017). Este trabajo define seis tipos de perfiles (Observador Ocupado, Explorador Equipado, Enlace Local, Intérprete Influyente, Mapeador Meticuloso y Campeón de la Comunidad) en función de su nivel de entendimiento de los datos abiertos, su actitud hacia ellos y su capacidad de influencia en la comunidad, y los caracteriza por sus habilidades en el uso de la tecnología, la recopilación y análisis de datos, la interpretación de información georreferenciada y el uso de los medios de comunicación. Los resultados de esta segmentación facilitan adaptar los servicios del portal, la oferta de servicios específica y la comunicación de todo ello a los distintos colectivos definidos.

A partir de este marco metodológico, y de forma alineada con la vocación establecida en la Estrategia Aragón Open Data, se definieron ocho tipos de colectivos de usuarios en el caso de Aragón, tal y como puede verse en la Figura 4.

Durante la fase inicial del proyecto, se trabajó en identificar el conocimiento y uso que los distintos colectivos identificados daban a las funcionalidades de los datos (informativo, consulta, explotación, descarga y visualización) del portal Aragón Open Data. De manera adicional, cada colectivo se caracterizó con base en una serie de atributos

(entre ellos, las funcionalidades citadas para las cuales utilizaban la información y datos disponibles), caracterizando cada recurso de información disponible por estos atributos a su vez, dando lugar a un cruce de información que permitió evaluar la cantidad de información potencialmente reutilizable por cada uno de los colectivos, según los parámetros definidos. Los resultados de este trabajo se resumen en la Tabla II.



Figura 4. Colectivos de usuarios definidos en el marco del proyecto y correspondencia con los colectivos contemplados en la Estrategia Aragón Open Data 2019-2022

Colectivos de usuarios	Datasets con potencial de uso	% de potencialidad de uso
Narradores de historias	2.621	88%
Desarrolladores y programadores	1.955	66%
Empresas reutilizadoras de datos	1.561	53%
Empleados de las Administraciones públicas	2.735	92%
Dirección, gestores y altos cargos	620	21%
Fiscalizadores de la acción pública	647	22%
Ciudadanía	647	22%
Estudiantes	422	14%

Tabla II. Colectivos de usuarios, conjuntos de datos disponibles por colectivo y potencialidad de uso

Una vez realizado este diagnóstico, y con el fin de dar respuesta de la forma más adecuada a las necesidades de los usuarios, se inició un proceso de contraste con los distintos definidos. Para ello, se organizaron distintas mesas de trabajo temáticas en las que se establecieron debates en torno a los datos abiertos y se indagó en las expectativas y necesidades de estos distintos colectivos definidos.

Como consecuencia de este proceso de contraste se extrajeron una serie de conclusiones relativas a la accesibilidad de los datos, su reutilización y utilidad, para poder determinar si los mismos eran adecuados para dar respuesta a las necesidades reales de cada colectivo. En concreto, se observó la necesidad de fomentar la reutilización de datos por los distintos grupos de interés o colectivos de usuarios, formar a la comunidad de usuarios en materia de Open Data para que puedan extraer el mayor valor añadido posible de la plataforma, facilitar el uso y entendimiento del portal, e impulsar una comunidad de usuarios que interactúen y compartan sus conocimientos.

## 5. Resultados

### 5.1. Arquitectura y funcionalidad del nuevo servicio

A partir del trabajo realizado en la primera fase, comenzó a desarrollarse la herramienta que permitiría la creación de historias por parte de los usuarios. Dichas historias se basarán en datos disponibles en el portal de Open Data del Gobierno de Aragón, de manera que los usuarios puedan construir un relato atractivo y diferencial en torno a un conjunto de datos disponible, otorgándoles un contexto y una interpretación que pueda ser de interés para el resto.

En este sentido, la plataforma, permitirá a los usuarios observar y leer las historias generadas por otros y realizar aportaciones o contribuciones a las mismas, de manera que se fomente un espacio de co-creación y colaboración entre distintos usuarios. El objetivo, será el de crear un portal que albergue de manera progresiva las historias generadas por los distintos grupos y colectivos de usuarios, contadas a través de datos abiertos.

La implementación del nuevo servicio se ha llevado a cabo en la arquitectura actual de entornos de la plataforma Aragón Open Data, si bien se han producido modificaciones en algunos de los componentes para la configuración de Aragón Open Data Focus:

1. Open Data Focus: La tecnología utilizada para el frontend de Open Data Focus es Angular. Reutilizando el proyecto base de Open Visual Data, se ha creado un servicio como un portal nuevo, dejando el anterior servicio como mera herramienta.
2. Open Visual Data: El proyecto de Open Visual Data, con ese mismo nombre, quedará discontinuado en su repositorio de código. Mientras tanto, la nueva herramienta de creación de gráficos de Aragón Open Data vivirá dentro del nuevo proyecto Open Data Focus, utilizando la misma tecnología, Angular.
3. Focus API: El proyecto Open Data Focus necesita de una potente API que interactúe con la base de datos, para ello se reutiliza la API de Open Visual Data, desarrollada en Spring Boot y se nutre de nuevas funcionalidades, se mejora lo que ya había y se condiciona para cumplir los requisitos del proyecto.
4. Visual Data API: El proyecto Open Visual Data, contenía una API que quedará discontinuada en su repositorio de código.
5. Update Charts: Es un proceso NodeJS, que se ejecuta diariamente y actualiza todas las gráficas disponibles en la plataforma.
6. Base de datos: Utilizando la tecnología MongoDB, se mantiene la misma base de datos que se utilizaba para el proyecto de Open Visual Data. Tan solo sufre algunos cambios de estructura debido a las nuevas funcionalidades y al proceso interno de actualización diario de las gráficas.

Con el objetivo de mantener la homogeneidad en la infraestructura del proyecto de Aragón Open Data, las tecnologías utilizadas para este desarrollo han sido NodeJS para el backend y servicios web, Angular para la parte frontal y visual del portal, y PostgreSQL como soporte de almacenamiento de datos.

A nivel de funcionalidades, el usuario será capaz de interactuar con la parte visual de Aragón Open Data Focus (Angular), gracias a un navegador web. Todas las peticiones que realice el usuario serán recogidas a través de la lógica de Angular, que llamará a su vez a la API localizada en el backend (NodeJS) en continuo estado de escucha. La API (NodeJS) se encargará de transformar estas peticiones del usuario en consultas a la base de datos (PostgreSQL), que devolverá los resultados pedidos con el objetivo de que el frontal (Angular) sea capaz de mostrarlos en forma de historia.

De cara a la utilización del servicio, se ha llevado a cabo un esfuerzo específico para garantizar un

uso lo más sencillo e intuitivo posible, con una interfaz única independientemente del tipo de usuario que haga uso del servicio, al que no se le solicita ningún dato personal para poder utilizar el servicio, en línea con los principios rectores de los datos abiertos que propugnan una accesibilidad total de los mismos. Al acceder a Aragón Open Data Focus, el usuario se encuentra ante una pantalla dividida en cuatro secciones: título de la historia, descripción (con un máximo de 420 caracteres), categorización (que permite añadir una o varias categorías a la historia para facilitar la búsqueda de textos relacionados en función de las temáticas definidas) y secciones. Puede verse una de las funcionalidades descritas en la Figura 5.

Figura 5. Funcionalidad “Añadir sección”

En cada una de las secciones, el usuario de Aragón Open Data puede incluir distintos recursos que le permitan reforzar los mensajes que compongan su texto: desde vídeos de la plataforma YouTube hasta presentaciones de SlideShare o imágenes publicadas en Internet. En esta funcionalidad es donde destaca la posibilidad de crear gráficos a partir de los recursos del banco de datos de Aragón Open Data, ya sean visualizaciones preexistentes u otras nuevas a partir de los datos abiertos del portal.

Así, el proceso de co-creación queda configurado de la siguiente manera: el Gobierno de Aragón publica y pone a disposición de cualquier usuario los datos a través del portal Aragón Open Data, y provee este servicio que permite el tratamiento de los datos de forma sencilla a través de la propia web, siendo los usuarios los que deciden cómo explotan estos datos para respaldar sus propias historias. Adicionalmente, y como ya se ha indicado, si bien se ha trabajado en que el servicio web sea lo más intuitivo posible, también se ha creado un vídeo tutorial (8), accesible a través de la web y alojado en YouTube, para facilitar la comprensión y uso del servicio, si bien también pueden realizarse preguntas más concretas a través del correo electrónico [opendata@aragon.es](mailto:opendata@aragon.es). Una vez finalizada la historia, es moderada por el propio equipo responsable del portal, si bien esta mo-

deración se limita a la corrección gramatical y ortográfica de los textos presentados y a que el contenido no sea inadecuado (por ejemplo, referido a actividades ilícitas, como el racismo, la xenofobia o la apología del terrorismo, o nocivos, peligrosos o poco saludables) y no infrinja las leyes de protección de datos o propiedad intelectual.

Adicionalmente, y con el objetivo de capacitar a los usuarios y dar a conocer la plataforma, se ha creado una formación ad hoc sobre Aragón Open Data Focus y cómo contar historias con datos que capacitará a los empleados de la Administración pública en esta primera fase del despliegue del nuevo servicio.

## 5.2. Aragón Open Data Focus: las historias al servicio de los datos

Como consecuencia de los trabajos realizados con los distintos colectivos de usuarios consultados, se identificaron mejoras diversas a realizar en el portal de Aragón Open Data. Adicionalmente, en relación con el servicio Aragón Open Data Focus (9), se perfilaron diversas temáticas acerca de las cuales sería de interés para los usuarios que se escribiera una historia, como, por ejemplo, los alojamientos turísticos rurales disponibles en Aragón, un análisis del gasto que se destina a la educación en el territorio o el avance de Aragón en su transformación hacia una sociedad de la información. Así, con el servicio recién estrenado, ya están disponibles algunas historias acerca de las temáticas consideradas de interés. A modo de ejemplo se presentan las siguientes:

- Destaca en primer lugar, por su sentido de actualidad, la historia “El coronavirus en Aragón” que relata la incidencia de la COVID-19 en el territorio y su evolución. La historia cuenta, basándose en los datos disponibles, cómo Zaragoza ha sido la más damnificada de las provincias de la comunidad autónoma, concentrando 6 de cada 10 casos confirmados en la región.

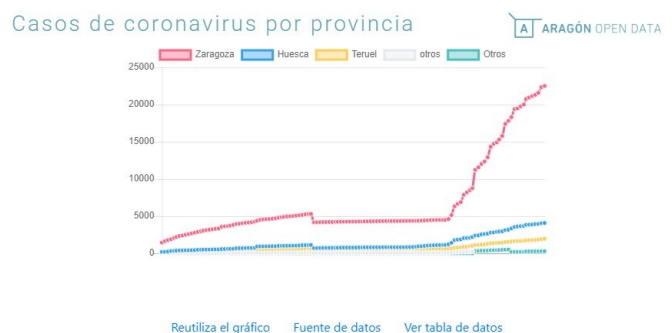


Figura 6. Historia “El coronavirus en Aragón”



Asimismo, la historia ofrece la visualización gráfica de la curva de la pandemia en el territorio y las distintas medidas que van a ser adoptadas para paliar las consecuencias, así como los fondos económicos destinados a la reconstrucción de la región o el rol del sistema aragonés de salud en la lucha contra la pandemia. Un ejemplo visual de los gráficos creados para esta historia puede consultarse en la Figura 6.

- Asimismo, destaca la historia que analiza el impacto de la COVID-19 en el turismo aragonés, “Este verano salimos juntos en Aragón”, que trata de poner en valor el atractivo paisajístico, cultural y gastronómico que ofrece el territorio aragonés basándose en datos e información pública, tratando de incentivar el turismo interior en una época en que se recomienda evitar los desplazamientos en la medida de lo posible. La historia analiza los datos relativos a la oferta de plazas hoteleras y cuenta los atractivos culturales existentes en el entorno de Zaragoza y las oportunidades de turismo activo o rural y gastronómico que ofrece el entorno de Huesca, con el objetivo de atraer turistas locales y de otras comunidades autónomas.



Figura 7. Historia “Este verano salimos juntos en Aragón”

- Otra de las historias destaca el impacto de la COVID-19 en el colectivo universitario del territorio, y, específicamente, cómo la pandemia ha afectado a la movilidad de los estudiantes universitarios. Basándose en los datos históricos relativos a los acuerdos Erasmus y SI-CUE de los estudiantes de la Universidad de Zaragoza, se analizan los destinos favoritos de los estudiantes, resultando como destinos más atractivos Granada, a nivel nacional, e Italia, a nivel internacional, pudiendo verse afectada la tendencia de cursar estudios fuera en el presente curso. Asimismo, la historia trata de relatar cómo ha evolucionado la solicitud para cursar determinados estudios por parte del alumnado, habiendo ganado relevancia todo lo relativo a los grados y titulaciones de la rama biosanitaria y los grados de

carácter más técnico, como matemáticas o física. Por último, la historia pone en contexto el dato relativo a las ayudas o becas a estudiantes, reforzando la idea de que la COVID-19 no puede ser un impedimento para el desarrollo formativo adecuado de la juventud aragonesa.



Figura 8. Historia “Un año universitario atípico”

- Además de las historias inspiradas por la pandemia, se han dibujado otras, como la historia “El avance de Aragón en su transformación hacia una sociedad de la información”, que trata de relatar los avances que se han dado en los últimos años en materia TIC y cómo estos han contribuido a la transformación de la sociedad aragonesa. El artículo pone de manifiesto el contexto caracterizado por la progresiva importancia de las TICs en las distintas vertientes de la vida, ocupando un rol fundamental en la mejora de la competitividad del territorio. En este contexto, se habla del esfuerzo conjunto realizado por parte del Gobierno, empresas y sociedad de Aragón para convertirse en una de las denominadas “regiones inteligentes”, utilizando para ello los datos relativos a la incorporación de profesionales expertos en las TIC en las empresas del territorio o la disponibilidad de ordenadores en determinados sectores de actividad.

## 6. Conclusiones

El nuevo servicio del proyecto de Aragón Open Data, Aragón Open Data Focus, se constituye como una experiencia innovadora y pionera a nivel estatal en materia de reutilización de datos del sector público. Este servicio permite a los usuarios del portal utilizar los datos abiertos del catálogo de Aragón Open Data para elaborar sus propias historias basadas en ellos y construir visualizaciones que faciliten o enriquezcan la comprensión de los lectores de la información del sector público.

Si bien aun es pronto para poder medir el éxito del proyecto, que vendrá determinado por el aumento del uso y reutilización de los datos y la mejora de la capacitación de los distintos colectivos de usuarios para reutilizar la información pública, pueden ponerse en valor algunos indicios alenta-

dores del impacto buscado por el equipo de la Dirección General de Administración Electrónica y Sociedad de la Información de Gobierno de Aragón: por un lado, destaca la adhesión e interés que ha suscitado este proyecto durante su desarrollo, con casi 300 participantes asistiendo a los distintos eventos online celebrados en los meses de mayo a septiembre de 2020; por otro, el uso que se está haciendo del servicio, con 14 historias publicadas a mayo de 2021. Estas primeras historias se espera que se conviertan en la primera piedra del proyecto de construcción de una comunidad empoderada de usuarios, activa y comprometida, que alimente el portal web con nuevas narrativas basadas en los datos abiertos del territorio, que sea capaz de generar un diálogo en torno a estos datos y que permita avanzar, a la postre, hacia un modelo de reutilización más intensiva de los datos abiertos de Aragón.

Así, a futuro será muy relevante que los datos disponibles sigan creciendo y mejorando su calidad para que los usuarios puedan disponer de información relevante, veraz y actualizada, y que las historias que se cuelguen en la plataforma puedan cambiar y evolucionar a la vez que los datos. La actualización de los datos contribuirá a nuevas interpretaciones y relatos vivos sobre distintos ámbitos y cuestiones que afectan al territorio, dando un valor diferencial a la información, ya que los datos en sí mismos, sin contexto e interpretación, carecen de valor.

## Notas

- (1) <https://opendata.aragon.es/>
- (2) <https://opendata.aragon.es/datos/catalogo>
- (3) <https://datausa.io/>
- (4) <https://es.datachile.io/>
- (5) <https://dataafrica.io/>
- (6) <https://www.ers.usda.gov/data-products/data-visualizations/>
- (7) <https://data.gov.sg/>
- (8) [https://www.youtube.com/watch?v=AnNnt6weWx4&ab\\_channel=ARAGONOPENDATA](https://www.youtube.com/watch?v=AnNnt6weWx4&ab_channel=ARAGONOPENDATA)
- (9) <https://opendata.aragon.es/servicios/focus/>

## Referencias

- Abella, Alberto. (2019). La reutilización de datos abiertos en España II. desdeDatum. [https://www.desidedatum.com/wp-content/uploads/2019/12/La\\_reutilizacio%CC%81n\\_datos\\_abiertos\\_en\\_espan%CC%83a\\_2019.pdf](https://www.desidedatum.com/wp-content/uploads/2019/12/La_reutilizacio%CC%81n_datos_abiertos_en_espan%CC%83a_2019.pdf)
- Aleixandre-Benavent, Rafael; Ferrer Sapena, Antonia; Peset, Fernanda. (2019). Compartir los recursos útiles para la investigación: datos abiertos (open data). // Educación Médica, ISSN: 1575-1813. Elsevier. 22:2. DOI: 10.1016/j.edumed.2019.07.004

- Álvarez García, Sergio; Gértudix Barrio, Manuel; Rajas Fernández, Mario. (2014). La construcción colaborativa de bancos de datos abiertos como instrumento de empoderamiento ciudadano. // Revista Latina de Comunicación Social. 69:6, 661-683. Universidad de La Laguna. [http://www.revistalatinacs.org/069/paper/1029\\_UR/RL\\_CS\\_paper1029.pdf](http://www.revistalatinacs.org/069/paper/1029_UR/RL_CS_paper1029.pdf)
- Attard, Judie; Orlandi, Fabrizio; Scerri, Simon; Auer, Sören (2015). A systematic review of open government data initiatives. // Government Information Quarterly. 32:4, 399-418. <https://www.sciencedirect.com/science/article/abs/pii/S0740624X1500091X?via%3Dihub>
- Bichard, Jo-Anne; Knight, Gail (2012) Improving public services through open data: public toilets. // Municipal Engineer. 165:ME3, 157-165. <https://researchonline.rca.ac.uk/1119/1/Bichard%20%20Knight.pdf>
- Blanes Climent, Miguel Ángel (2013). La transparencia informativa de las administraciones públicas: el derecho de las personas a saber y la obligación de difundir información pública de forma activa. Universidad de Alicante. [http://rua.ua.es/dspace/bitstream/10045/35719/3/tesis\\_miguelangel\\_blanescliment.pdf](http://rua.ua.es/dspace/bitstream/10045/35719/3/tesis_miguelangel_blanescliment.pdf)
- Cabezali, Pablo. (2020) Diagnóstico de los datos abiertos en España, evaluación de los portales autonómicos de datos abiertos y desarrollo de cuadros de mando para la exposición de resultados. Universidad de Deusto. <https://dkh.deusto.es/comunidad/explorer/recurso/diagnostico-de-los-datos-abiertos-en-espa/54b38648-d605-4974-a441-7d869e020391>
- Caro-Castaño, Lucía; Selva-Ruiz, David (2020). Data storytelling: el empleo de datos abiertos en la construcción de relatos publicitarios de marca. // Comunicación y hombre: revista interdisciplinar de ciencias de la comunicación y humanidades. 16, 35-57. Universidad Francisco de Vitoria. <https://portalderevistas.ufv.es/index.php/comunicacionyhombre/article/view/572/585>
- Cerrillo i Martínez, Agustí (2011). Transparencia administrativa y lucha contra la corrupción en la Administración local. // Anuario del Gobierno Local. 1, 277-313. [http://repositorio.gobiernolocal.es/xmlui/bitstream/handle/10873/1255/16\\_Cerrillo\\_Transparencia.pdf?sequence=1&isAllowed=y](http://repositorio.gobiernolocal.es/xmlui/bitstream/handle/10873/1255/16_Cerrillo_Transparencia.pdf?sequence=1&isAllowed=y)
- Clabo, Néstor; Ramos-Vielba, Irene (2015). Reutilización de datos abiertos en la administración pública en España y uso de licencias-tipo. // Revista Española de Documentación Científica. 38:3, e097. <http://redc.revistas.csic.es/index.php/redc/article/view/900/1269>
- Conejero Paz, Enrique (2013). Gobierno abierto y democracia participativa. // 3c Empresa: investigación y pensamiento crítico. 2:4, 2013. <https://dialnet.unirioja.es/descarga/articulo/4817923.pdf>
- Cordero Sanz, Carlos (2007). Derecho de acceso a la información y transparencia: herramientas necesarias para la democracia en una sociedad abierta. // El derecho de acceso a la información pública, 55-67. [https://eprints.ucm.es/id/eprint/8946/1/MANUEL\\_SANCHEZ\\_DER\\_ACCESO\\_V14\\_2.pdf](https://eprints.ucm.es/id/eprint/8946/1/MANUEL_SANCHEZ_DER_ACCESO_V14_2.pdf)
- Cotino Hueso, Lorenzo (2015). La nueva Ley de transparencia y acceso a la información. // Anuario de la Facultad de Derecho. 241-256. [https://ebuah.uah.es/dspace/bitstream/handle/10017/22075/nueva\\_cotino\\_AFDUA\\_2014.pdf?sequence=1&isAllowed=y](https://ebuah.uah.es/dspace/bitstream/handle/10017/22075/nueva_cotino_AFDUA_2014.pdf?sequence=1&isAllowed=y)
- Dykes, Brent (2015). Data storytelling: What it is and how it can be used to effectively communicate analysis results. // Applied Marketing Analytics. 1, Issue 4. HSTalks. <https://hstalks.com/article/619/data-storytelling-what-it-is-and-how-it-can-be-use/>

- European Data Portal (2020). The Economic Impact of Open Data: Opportunities for value creation in Europe. ISBN: 978-92-78-42108-3. European Commission. <https://www.europeandataportal.eu/es/highlights/the-economic-impact-of-open-data>
- European Data Portal (2020). Analytical Report 12: Business-to-Government Data Sharing. ISBN: 978-92-78-41997-4. European Commission. [https://data.europa.eu/sites/default/files/analytical\\_report\\_12\\_business\\_government\\_data\\_sharing.pdf](https://data.europa.eu/sites/default/files/analytical_report_12_business_government_data_sharing.pdf)
- Flores Vivar, Jesús; Salinas, Aguilar, Cecilia. (2013). El periodismo de datos como especialización de las organizaciones de noticias en Internet. // *Correspondencias y Análisis*. 3, 15-34. <https://dialnet.unirioja.es/descarga/articulo/4739290.pdf>
- Gobierno de Aragón (2019). Estrategia Aragón Open Data [2019-2022]. [https://opendata.aragon.es/static/public/documentos/20190405\\_estrategia\\_opendata.pdf](https://opendata.aragon.es/static/public/documentos/20190405_estrategia_opendata.pdf)
- Greed, Clara (2003). *Inclusive Urban Design: Public Toilets*. Routledge.
- Kucera, Jan; Chlapek, Dusan (2014). Benefits and Risks of Open Government Data. // *Journal of Systems Integration*. ISSN: 1804-2724. 5:1. <http://si-journal.org/index.php/JSI/article/viewFile/185/254>
- Lee, Bongshing; Riche, Nathalie Henry; Isenberg, Petra; Carpendale, Sheelagh (2015). More than telling a story: transforming data into visually shared stories. // *IEEE Computer Graphics and Applications*. 35:5. <https://www.microsoft.com/en-us/research/uploads/prod/2016/12/StorytellingProcess-CGA2015.pdf>
- Montero Caro, María Dolores (2020). El encaje constitucional del modelo de Open Government. Universidad de Córdoba: UCOPress. <https://helvia.uco.es/xmlui/handle/10396/19485>
- NYC Open Data (2017). Meet the Users of Open Data. <https://thereboot.github.io/moti-personas/>
- Obama, Barack (2009). Transparency and Open Government Memorandum. <https://obamawhitehouse.archives.gov/the-press-office/transparency-and-open-government>
- Open Knowledge Foundation (2012). Open Data Handbook. <https://opendatahandbook.org/guide/es/what-is-open-data/>
- Palomares Herrera, Manuel (2017). Estudio comparado sobre transparencia y derecho de acceso en el ámbito internacional y su influencia en España. // *Ius Humani: Revista de Derecho*. 6, 123-153. <https://dialnet.unirioja.es/descarga/articulo/5877166.pdf>
- Ramos Simón, L. Fernando; Arquer Avilés, Rosario; Botezan, Iuliana; Cobo Serrano, Silvia; Sala Jiménez, Andrea; Sánchez Jiménez, Rodrigo; Del Valle Gastamiza; Félix (2012). De la reutilización de información del sector público a los portales de datos abiertos en Europa. // *BiD: Textos universitarios de biblioteconomía y documentación*. 29. <http://bid.ub.edu/29/ramos2.htm>
- Segel, Edward; Heer, Jeffrey (2010). Narrative Visualization: Telling Stories with Data. // *IEEE Transactions on Visualization and Computer Graphics*. 16:6. <https://ieeexplore.ieee.org/document/5613452>
- Sève, Bruno; Redondo, Ernesto (2019). El pabellón de deseos. Co-creación y co-instalación artística para la mejora del espacio público. *ACE: Architecture, City and Environment*. 14:42, 8200. <https://upcommons.upc.edu/bitstream/handle/2117/178902/8200-8741-1-PB.pdf?sequence=1>
- Suprata, Ferdian (2019). Data Storytelling with Dashboard: Accelerating Understanding through Data Visualization in Financial Technology Company Case Study. // *Journal Metris*. 20:1. <http://ojs.atmajaya.ac.id/index.php/metris/article/view/1441/1047>

---

Enviado: 2021-05-11. Segunda versión: 2021-05-31.  
Aceptado: 2021-06-16.

---



---

# A indexação social em catálogos digitais do *GoodReads* e *Skoob*

*La indexación social en los catálogos digitales de GoodReads y Skoob*

*Social indexing in GoodReads and Skoob digital catalogs*

---

Jessica Beatriz TOLARE, Mariângela Spotti Lopes FUJITA, Bruna Daniele de Oliveira SILVA

Universidade Estadual Paulista Júlio de Mesquita Filho, Campus de Marília, Dep. de Ciência da Informação, Av. Hygino Muzzi Filho, 737, Campus Universitário 17525900 - Marília, SP (Brasil), [jessica.tolare | mariangela.fujita | bruna.d.silva ]@unesp.br.

## Resumen

Se investigan los procedimientos y criterios de indexación social en las plataformas de libros de Skoob y GoodReads. Se realizó una investigación exploratoria dividida en investigación bibliográfica sobre indexación social y análisis de la atribución y recuperación de términos representativos en las plataformas de libros de Skoob y GoodReads. Los resultados de la investigación mostraron que ambas plataformas ofrecen los mismos servicios básicos: organizar las lecturas de los usuarios y permitir la interacción entre ellos con respecto a sus lecturas. Lo que las distingue son pequeños detalles, como que GoodReads no permite la indexación social, aunque los usuarios pueden asignar términos representativos a los libros que componen su estantería personal; mientras que, en Skoob, son los usuarios quienes realizan la indexación social. Ambas plataformas no cuentan con ningún tipo de vocabulario controlado, lo que afecta directamente a la representación y recuperación de libros.

**Palabras clave:** Indexación social. Representación y recuperación de información. Folksonomias. Skoob. Goodreads.

## 1. Introdução

A internet foi um marco histórico, mudando por completo a vida das pessoas. A sua expansão fomentou o surgimento de redes sociais digitais (Burlamaque & Barth, 2015).

Uma pesquisa mostrou que, no Brasil, 134 milhões de pessoas acessam a internet, no qual a maioria ocorre por celular (Agência Brasil, 2020). Outro estudo mostrou que 97,56% acessaram a internet em todos os dias no mês de junho de 2020, sendo que a grande maioria utilizou mais para entretenimento do que para o trabalho (E-Commerce Brasil, 2020). Um relatório da *Internet Advertising Bureau* Brasil (IAB – Brasil, 2014; 2020) mostrou que, em 2013, 30% dos sites visitados pela população brasileira era de redes sociais. Desde o começo da pandemia de Covid-19, em 2020, e a partir das pesquisas anteriores percebeu-se que o brasileiro passou duas vezes mais tempo na internet do que a média global.

## Abstract

The representation and retrieval of books on the Skoob and GoodReads platforms is explored. In order to investigate the procedures and criteria of social indexing in the book platforms of Skoob and GoodReads, exploratory research was carried out in two phases: bibliographical research on social indexing, and analysis of the attribution and retrieval of representative terms in the book platforms of Skoob and GoodReads. The research results showed that both platforms offer the same basic services: organizing the users' readings and enabling interaction between them regarding their readings. What sets them apart are small details: GoodReads does not allow social indexing, although users can assign representative terms to the books that make up their personal shelf; on Skoob, the users perform the social indexing. Both platforms do not have any type of controlled vocabulary, directly affecting the representation and retrieval of books.

**Keywords:** Social indexing. Information representation. Information retrieval. Folksonomy. Skoob. Goodreads.

As redes sociais são definidas por Recuero (2011), como um conjunto de dois elementos: os atores, que são os indivíduos sociais, podendo ser pessoas, instituições ou grupos. O segundo são os nós de uma rede e suas conexões, que são as interações (ou laços sociais) ocorridos entre os indivíduos. A popularização das redes sociais, segundo Amaral e Salvador (2018), trouxe tendências à segmentação em nichos de temas, como, por exemplo, sites voltados para a organização de livros, que tem como público-alvo leitores. Para isso, foram criados aplicativos e plataformas digitais como o Skoob, *GoodReads*, *Scribd*, *Anobii* e *Librarything*.

O Skoob, *GoodReads*, *Anobii* e *Librarything* são sites cujo o objetivo é organizar a estante virtual do leitor. Essas plataformas são gratuitas e possibilitam que os usuários adicionem livros, cataloguem, atribua avaliações, escreva resenhas, interajam com outros usuários e participem de fóruns de discussões. Diferente do *Scribd*, uma

plataforma paga, que disponibiliza acesso a *ebooks* e *audiobooks* por um valor mensal.

O processo de colaboração na representação, organização e compartilhamento de informação na internet ocorre através de uma relação implícita entre os usuários, estabelecendo uma metalinguagem compreendida e construída por todos os participantes. São atribuídas etiquetas ou palavras-chave com o intuito de realizar representações que possibilitem a recuperação da informação, constituindo, dessa forma, uma indexação social (Guedes, 2010; Gonçalves & Assis, 2016).

Sousa e Benetti (2016) desenvolveu um estudo cujo objetivo foi de explorar a indexação social realizada no Skoob pela análise dos conceitos utilizados em *tags*, comparando-os aos conceitos presentes em resenhas escritas pelos usuários. Os resultados mostraram que houve discrepâncias entre as *tags* no singular e plural e que os conceitos presentes nas *tags* e nas resenhas, de um modo geral, foram diferentes, podendo gerar problemas na busca e recuperação dos livros na plataforma.

Em outra pesquisa, Kipp, Beak e Graf (2015) coletaram *tags* atribuídas aos 100 livros proibidos pela *American Library Association* (ALA) nas plataformas do *Librarything*, *GoodReads* e *BiblioCommons* e analisaram os termos em busca de encontrar os motivos da censura. Os resultados demonstraram consciência dos motivos da censura, bem como falta de controle de vocabulário nas *tags* em forma de frases longas e curtas, variedade de termos pessoais ou que não representava o assunto substancialmente. Os autores recomendaram mais cuidados quanto ao uso de *tags*, pois as plataformas podem estar suscetíveis à “*trolling*”, ou seja, desorganizações e erros deliberados que causam bagunça dentro do ambiente na internet.

Vaidya e Harinarayana (2016, p. 289) contextualizam explicando que a atribuição de *tags* sociais aos objetos do conhecimento exigiu dos usuários uma nova abordagem na descoberta e uso de recursos. Nesse sentido, as autoras complementam que os bibliotecários estão familiarizados com diferentes ferramentas de organização do conhecimento, como esquemas de classificação e taxonomias, enquanto as *tags sociais* (também chamadas de folksonomia) são metadados gerados pelo usuário para recursos da web, principalmente, para descrever conteúdos de assuntos e ser usado para categorização de conteúdo e sua recuperação subsequente.

As redes sociais popularizaram e redefiniram a prática de etiquetagem social, com a possibilidade de criar *hiperlinks* instantaneamente. As *hashtags*

(#) são um tipo de etiqueta que classifica o conteúdo gerado nesses ambientes, criadas com o propósito de conectar tópicos, assuntos, produtos, mídias, etc., independente da conexão entre os usuários (Chang & Iyer, 2012). O uso da folksonomia para indexar conteúdos em ambientes digitais requer a necessidade de utilizar ferramentas como a sugestão de etiquetas e autopreenchimentos, proporcionando algum tipo de controle ortográfico e de sinônimo (Holstrom, 2018).

Da mesma forma que uma biblioteca tradicional ou digital precisa que seu acervo passe por uma organização de suas informações é necessário que essas plataformas na internet recebam o mesmo tratamento, pois trabalham com a organização de acervo pessoal em âmbito digital.

Nesse sentido, de acordo com estudos citados (Sousa & Benetti, 2016; Kipp, Beak & Graf, 2015), análise de relatos de usuários na internet e, após um período de uso do Skoob e do *GoodReads* de seis anos, desde 2015, percebeu-se a existência de problemas relacionados a busca e a recuperação de livros nessas plataformas. Nesse sentido, questionamos como a *indexação social de livros é realizada nos catálogos do Skoob e GoodReads e quais são os critérios para realizar esse processo*.

À vista disso, a pesquisa tem como objetivo analisar o processo e critérios de indexação social nas plataformas do Skoob e *GoodReads*. Para isso, foram realizadas pesquisa bibliográfica sobre indexação social e análise de como os termos representativos são atribuídos e recuperados nas plataformas do Skoob e *GoodReads* a partir das buscas dos livros que compõem os catálogos, verificando os critérios utilizados para realizar o processo de indexação. Há necessidade em se desenvolver o estudo devido ao uso crescente de usuários utilizando a internet para organizar o seu acervo pessoal e no interesse em obter interações com outros usuários.

## 2. A indexação social na internet

O uso e atribuição de termos representativos por usuários na literatura ainda está em desenvolvimento. Esse processo recebeu diferentes nomenclaturas de diversos autores, mas o seu significado continua sendo o mesmo.

Santos e Corrêa (2019, p. 6) apresentam alguns termos utilizados para nomear esse assunto: “classificação social”, “*social rank*”, “etiquetagem colaborativa”, “*social tagging*”, “representação colaborativa da informação”, “*collaborative representation of information*”, “indexação social”, “*social indexing*”, “modelos de colaboração”, “*collaboration models*”, “modelos colaborativos”, “*col-*

*laborative models*”, “modelos de indexação”, “*indexing models*”, “modelos colaborativos de indexação”, “*collaborative indexing models*” (Santos & Corrêa, 2019, p.6). Gómez-Díaz (2012, p. 103) define indexação social como um sistema de representação do conteúdo, em que os próprios usuários descrevem os recursos e compartilham essas representações por intermédio da web social com base na linguagem natural.

De acordo com Hassan-Montero (2006), a indexação social só deve ser utilizada quando os sistemas permitirem que vários usuários indexem um mesmo recurso, tornando o resultado da indexação uma descrição intersubjetiva e efetivando contratos semânticos por meio da relação de significados que surgem da indexação e que são incorporados pela linguagem em uso.

Guedes, Moura e Dias (2011) explicam que a indexação social quando considerada como uma modalidade da indexação tradicional é caracterizada pelo uso da linguagem natural e reconhecida por folksonomia, no âmbito das plataformas digitais colaborativas.

Segundo Gonçalves e Assis (2016), a indexação social possibilita uma modalidade de linguagem constituída de forma dinâmica e colaborativa, observando que há atuação de uma metalinguagem no processo de representação da informação por meio do uso de etiquetas (palavras-chave) constituindo uma folksonomia, no qual sua função é de estabelecer uma metalinguagem coletivamente construída no ambiente digital.

A Folksonomia é um neologismo (*folk+taxonomy* = povo + taxonomia) criado por Thomas Vander Wal, com o intuito de nomear o resultado de ações de representação da informação guiada pelos usuários em ambientes colaborativos mantidos por *softwares* sociais na *web*. Elas são consideradas como sistemas orgânicos, resultados da atribuição livre e pessoal de marcadores (palavras-chave) às informações ou objetos visando à organização e recuperação da informação (Vander Wal, 2015).

Vander Wal (2015) apresenta duas formas de construir a folksonomia: a primeira é a restrita, em que apenas um usuário ou um grupo pré-definido de usuários atribuem etiquetas a um único objeto e a segunda é a ampla, que ocorre em ambientes em que qualquer usuário pode atribuir etiquetas ao mesmo objeto.

Sterling (2005) define a folksonomia como uma metodologia de recuperação da informação espontânea e baseada na internet, que consiste em rótulos ou *tags* atribuídas de formas colaborativas, que categorizam conteúdos como recursos na *web*.

De acordo com Shirky (2005), as folksonomias são úteis para organizar recursos da informação e apoiar a recuperação eficiente de recursos. A atribuição de *tags* sociais oferecerá uma indexação baseada em assuntos, principalmente, em ambientes em que a indexação é cara, seja por causa do tamanho da coleção ou pela falta de estrutura.

A atribuição de etiquetas (*tags*), possibilitada pela interação entre os usuários, pode ou não ser consensual. Essa ação é marcada por negociações e pactos, existindo a possibilidade de se apoiarem ou se contraporem parcialmente ou totalmente (Fiorin, 2006; Guedes, Moura & Dias, 2011).

Guedes, Moura e Dias (2011) defendem que a folksonomia possui um caráter polifônico, visto que, sob a abordagem da organização da informação, ela possibilita um diálogo enunciativo, ao dar condições iguais aos usuários que fazem atribuição do sentido às etiquetas. Entretanto, para Assis e Moura (2013, p. 86), deve ocorrer a linguagem mista (controlada e livre), uma alternância entre o uso do vocabulário e o emprego da linguagem natural, assim como na interoperabilidade entre essas linguagens, com o intuito de construir modelos e serviços voltados para a organização da informação na *web* social.

Dziekaniak, Pacheco e Kern (2011) e Barros (2011) ressaltam a importância da coexistência de linguagens de indexação controladas e livres. Santos e Corrêa (2015) complementam explicando que o vocabulário livre utilizado pelos usuários deve contribuir para o aprimoramento da linguagem de indexação controlada. Dessa forma, Amaral e Salvador (2018) explicam que em um sistema folksonômico como processo é caracterizado pelo cruzamento de *tags* oficialmente definidas pelos bibliotecários com as *tags* atribuídas pelos usuários a partir da escolha de um gênero literário.

Kroski (2005) resalta a importância da coexistência das linguagens ao explicar que a folksonomia possui problemas, carecendo de precisão na recuperação da informação quando não há o uso de vocabulário controlado. Esse problema se torna mais explícito ao ser evidenciado pela falta de controle de sinônimos e hierarquias e pela possibilidade de ocorrer manipulações com o intuito de tornar populares as *tags* que mais interessam aos usuários. Vaidya e Harinarayana (2016) complementam que o uso de *tags* permite que os usuários busquem facilmente as informações que desejam, através do uso de termos em comum e sem se preocupar com a complexidade dos mecanismos do sistema de indexação.

A indexação social é utilizada para definir o processo de atribuição de termos representativos realizado por usuários na internet que gera vocabulários livres, denominados folksonomias, para auxiliar o usuário no uso de um ambiente virtual. Muitos sites utilizam a colaboração dos usuários para realizar a indexação social. Para entender como funciona esse processo e como afeta a representação e recuperação da informação é preciso analisar a indexação social dessas plataformas. Nos procedimentos metodológicos e na discussão dos resultados é apresentado como foram realizadas as análises propostas no estudo e os resultados obtidos.

### 3. Procedimentos metodológicos

A pesquisa exploratória e analítica, de cunho qualitativo realizou análise qualitativa da coleta dos dados da representação e recuperação de livros nas plataformas digitais de organização de livros, *GoodReads* e *Skoob*, a partir de simulações de buscas.

O catálogo digital de livros *GoodReads* é um site lançado em janeiro de 2007 e comprado pela Amazon em 2013. Seu cadastro é gratuito e é possível adicionar livros a suas estantes, avaliar e ver o que os outros usuários estão lendo. O idioma oficial é o inglês, mas muito livros são cadastrados por usuários em português.

O *Skoob*, um anagrama da palavra *books* (livros), é um catálogo digital brasileiro de compartilhamento de informações sobre leitura criado por Lindenberg Moreira em 2009. Ele é alimentado exclusivamente por usuários cadastrados (leitores, autores e editoras) e funciona como uma estante virtual para auxiliar os usuários com a organização de suas leituras (*Skoob*, 2021).

Ambas as plataformas não disponibilizam acesso ao conteúdo dos livros. Apenas as informações técnicas e resumo são disponibilizados com o objetivo de organizar, separando o acervo por livros "Lidos", "Não lidos", "Favoritos", "Abandonados" e entre outros.

O livro escolhido para realizar as simulações de buscas foi "O mundo Assombrado pelos demônios", de Carl Sagan, um livro científico que trata sobre assuntos como ciências e pseudociências. O objetivo é observar como ele está sendo representado e recuperado nas plataformas do *GoodReads* e do *Skoob*.

### 4. Discussão dos resultados

Baseando-se no quadro desenvolvido por González-Fernández-Villavicencio (2019, p. 146-148), em que compara os sistemas do *GoodReads* e do *Librarything*, foi criado o Quadro 1 a fim de comparar os sistemas do *Skoob* e do *GoodReads*.

Características	<i>Skoob</i>	<i>GoodReads</i>
Descrição	Rede social para leitores do Brasil. Estante virtual para organização de livros. "Socializar e incentivar o hábito da leitura" ( <i>Skoob</i> , 2021).	Rede social para discussão da leitura Focado em livros e atividades relacionadas a leituras e interação entre usuários.
Modalidade	Assíncrona	Assíncrona
Ano de criação	2009	2006
Catálogo	Não disponível	1,5 bilhão de livros 50 milhões de resenhas
Usuários	7 milhões de usuários	55 Milhões de usuários
Escopo temático	Literatura em geral	Literatura em geral
Fundadores	Lindenberg Moreira	Otis Chandler e Elizabeth K. Chandler. Comprado pela Amazon em 2013.
Tecnologia	Blogs e redes sociais	Blogs e redes sociais.
Recursos e serviços oferecidos	Avaliação através de atribuição de estrelas (1 a 5), comentários dos livros e interação com leitores, autores e editoras. Clube de assinatura, que, por um valor mensal, o usuário recebe em casa uma caixa com livros e brindes, todos relacionados a leitura. Permite que qualquer usuário, independentemente de sua modalidade (autor, leitor ou editora), possa realizar o cadastramento dos livros na plataforma	Possibilita a avaliação e comentários sobre livros, permitindo a interação com outros leitores e autores, recebendo recomendações com base nas classificações das avaliações dos livros adicionados. Permite a importação e exportação de registros bibliográficos. Mas os o sistema de importação de livros que não estão em seu catálogo é complicado e desajeitado.



<b>Modos de participação</b>	Possibilita que os usuários troquem impressões sobre livros, através de fóruns de discussões e resenhas avaliativas. Classificar os livros de acordo de acordo com a experiência de leitura. Os usuários podem atribuir tags e gêneros representativos aos livros no registro geral do livro e especificamente dentro da sua estante	Recomenda o livro com base nas informações personalizadas sobre um leitor na rede (resenhas, avaliações, autores, etc.). Oferece grupos de discussão para comentários de livros (clubes do livro, grupo de estudos, etc.). Traz jogos relacionados ao mundo do livro e fornece reportagens sobre novas publicações que são complementadas com rankings de romances, entrevistas com autores e até prêmios literários
<b>Ferramentas de participação</b>	Existe a meta de leitura, criada pela plataforma, com o objetivo de incentivar o uso da leitura e da plataforma. A cada 250 páginas lidas e registradas, o usuário ganha uma medalha em seu perfil. Existem sorteios de cortesias, em que as plataformas, em parceria com editoras, sorteiam livros para os usuários. Na modalidade disponível para iOS e Android permitem que os usuários acessem To be read (TBR), em que a plataforma sorteia um dos livros da meta de leitura do usuário para ser a próxima leitura dele. Possui um sistema de troca de livros em âmbito nacional, onde os usuários podem realizar trocas de livros. Alguns livros possuem links para compras de diferentes lojas online.	Publica boletins mensais de novos lançamentos e divulga trabalhos dos autores. Oferece um espaço que permite aos usuários fazer listas dinâmicas para indicar as obras que leram, dar-lhes uma avaliação e indicar leituras futuras. Produz uma seleção anual dos melhores livros do ano com base nas recomendações dos leitores chamada GoodReads Choice Awards. Organiza regularmente encontros do mundo real para trocar livros ou para visitas a pubs literários. Fornece links para compra de livros na Amazon.

Tabela 1. Comparação entre os sistemas do Skoob e do GoodReads

De acordo com a Tabela 1, é possível observar as informações técnicas, serviços oferecidos, participação dos usuários, ferramentas e recursos do Skoob e do *GoodReads*. Ambas as plataformas possuem como objetivo básico oferecer os mesmos serviços: possibilitar o usuário de organizar as suas leituras e interagir com os outros usuários sobre suas experiências com elas. Mas quando comparada mais detalhadamente fica nítido que o *GoodReads* é uma plataforma maior do que o Skoob, com mais recursos, livros e usuários cadastrados, podendo ser considerada a plataforma mais utilizada mundialmente, chegando ao ponto de oferecer prêmios em dinheiro.

Em contrapartida, o Skoob possui serviços e recursos mais específicos como clube de assinatura, troca de livros, sorteio de próximas leituras, sorteio de cortesias e entre outros. Um recurso disponível e muito utilizado é o acesso do usuário direto no registro geral do livro, podendo realizar alterações, atribuir tags e gêneros literários, mas sem nenhum tipo de critério, diretriz ou controle de vocabulário para preencher as informações.

A análise dos resultados foi dividida em duas seções: 4.1 – Análise dos procedimentos e critérios de indexação social e 4.2 – Análise da busca e recuperação de livros nos catálogos digitais.

#### 4.1. Análise dos procedimentos e critérios de indexação social

No final da página do perfil do livro “O mundo assombrado pelos demônios”, de Carl Sagan, no Skoob, fica disponível a possibilidade do usuário acessar os metadados do livro e editar as suas

informações. Acessando os registros, até o momento, a página apresenta pelo menos 12 registros da mesma edição, em que há muitos registros duplicados, que podem gerar problemas para a representação e recuperação das informações (Figura 1). Esse tipo de indexação social é caracterizado como ampla, pois qualquer usuário pode atribuir etiquetas ao mesmo objeto, sem ter nenhuma restrição, guia ou manual de como fazer o processo (Vander Wal, 2015).

Figura 1. Metadados do livro “O mundo assombrado pelos demônios”

Na Figura 1, são apresentados os metadados do livro “O mundo assombrado pelos demônios”. Logo acima, o Skoob ressalta a importância de as informações serem preenchidas corretamente e, que em caso de dúvida, aconselha a procurar as informações em sites de livrarias ou no próprio Google, mas não apresenta nenhuma diretriz de como preencher as informações ou quais consideram essenciais, o que pode causar problemas para o registro.

No final da página (Figura 1), antes de salvar o registro, há um campo para preencher com *tags*, como sugestão de inserir mais palavras-chave, que ajudarão a encontrar o livro na busca. Analisando o registro do livro “O mundo assombrado pelos demônios” é possível observar que não foi inserida nenhuma *tag*. No registro foram atribuídos os seguintes gêneros literários: “Filosofia”, “Não-ficção” e “Sociologia”. Esses gêneros não dão nenhum tipo de cobertura específica ou exaustiva para representar a essência da obra. Apesar disso, o site mostra uma lista simples de gêneros literários que são utilizados como termos representativos (Figura 2).

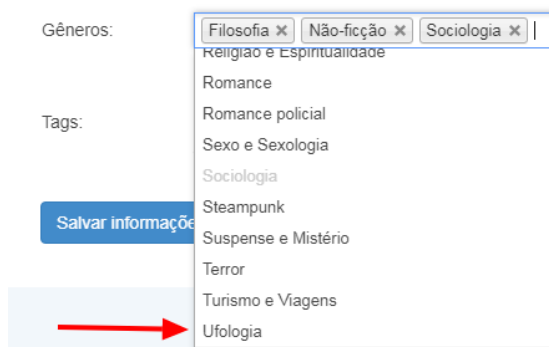


Figura 2. Lista de gêneros literários como termos representativos

Na Figura 2, é apresentada uma lista de gêneros literários utilizados para representar o assunto do livro no registro. Ao analisar mais detalhadamente essa lista de termos pré-definidos, pode-se observar que são poucos assuntos retratados, faltando muitas áreas do conhecimento e, os poucos termos que compõem a lista, não abordam completamente o assunto do livro. Nesse ponto, Gómez-Díaz (2012, p 103), considera a indexação social como um sistema de representação de conteúdo, no qual os próprios usuários descrevem os recursos e compartilham essas representações através da *web* social se baseando em uma linguagem natural, utilizada no cotidiano.

No registro do livro analisado, a lista não tem a possibilidade de acrescentar os termos “Ciência” ou “Pseudociência”, pois ambos os assuntos são

tratados no decorrer do texto. A plataforma não explica quem é o responsável pela criação da lista, não permite que usuários acrescentem mais termos e se há atualizações deles ou se são utilizados para a busca e recuperação do livro. A lista de termos apenas existe e não há nenhuma outra informação disponível a respeito dela, porém, o usuário é obrigado a preencher todas as informações no cadastramento do livro. O indicado seria, de acordo com Dziekaniak e Kern (2011), Barros (2011) e Santos e Corrêa (2015), que houvesse coexistência entre linguagens de indexação e vocabulário livre, sendo essa uma forma de complementar e contribuir para o aprimoramento da linguagem controlada e para a representação e recuperação da informação.

Enquanto que no GoodReads a indexação é realizada por um bibliotecário indexador. A única participação do usuário na indexação né que, na página do livro “O mundo assombrado pelos demônios”, existe uma lista com *tags* na lateral direita que direciona para outros livros do mesmo assunto. Não há nenhuma informação disponível a respeito do uso (ou não) de vocabulário controlado, mas, mesmo assim, os usuários podem atribuir *tags* para os livros compostos dentro da sua estante particular, facilitando a busca pelo livro posterior. Amaral e Salvador (2018) explicam que, uma vez atribuído o gênero literário, o próprio sistema cruza as *tags* oficiais definidas pelo bibliotecário com as *tags* escolhidas no sistema pelos usuários (Figura 3).

GENRES	
Science	3,580 users
Nonfiction	2,076 users
Philosophy	671 users
Religion	298 users
History	176 users
Philosophy > Skepticism	160 users
Religion > Atheism	137 users
Science > Popular Science	106 users
Science > Physics	87 users
Psychology	84 users

[See top shelves...](#)

Figura 3. Tags do livro ‘O mundo assombrado pelos Demônios’, Carl Sagan

Na Figura 3, é possível ver as *tags* atribuídas como termos representativos do livro “O mundo assombrado pelos demônios”, de Carl Sagan. A partir dessas categorias, é possível acessar outros livros que possuem as mesmas *tags* atribuídas e observar quantos usuários atribuíram esses mesmos termos especificamente dentro do seu acervo pessoal. Como exemplo, pode-se ob-

servar que o termo “*Science*” se repete, com especificações: “*Science > Popular Science*” e “*Science > Physics*”.

O uso de *tags* nos processos de representação da informação pelos usuários no *GoodReads* possibilita uma nova modalidade constituída de forma dinâmica e colaborativa, no qual estabelece uma metalinguagem coletivamente construída no ambiente digital (Gonçalves & Assis, 2016), pois, através dessas *tags*, é possível acessar a estante pessoal de outros usuários e obter informações a respeito de suas leituras e organizações do acervo.

#### 4.2. Análise da busca e recuperação de livros nos catálogos digitais

A partir das buscas realizadas nas plataformas foi possível obter alguns resultados. Foi realizada a pesquisa na barra de busca do *GoodReads* com e sem erros de digitação, a fim de observar a sua recuperação (Figura 4).

##### Search



Figura 4. Busca e recuperação do livro ‘O mundo assombrado pelos Demônios’, Carl Sagan na base de dados do GoodReads



Figura 5. Pesquisa na barra de busca do Skoob

A Figura 4, mostra a recuperação do livro “O mundo assombrado pelos demônios”, de Carl Sagan. Ambas as buscas obtiveram o mesmo resultado em que foi recuperado o livro. Nessa plataforma há um bibliotecário responsável pela criação e adoção de palavras-chave para descrever o conteúdo do livro a partir do seu gênero literário. A construção da folksonomia, ocorre de forma restrita, no qual apenas um usuário (o bibliotecário) atribui etiquetas a um único objeto (Vander Wal, 2015).

Na Figura 5 são apresentadas duas imagens em uma única figura: a primeira, que está na parte superior, consiste na busca pelo título do livro “O mundo assombrado pelos demônios”, sem erro de digitação. Pode-se observar que quando é realizada uma pesquisa, ela própria apresenta sugestões de livros que o usuário está ou pode estar procurando. Quando se observa a segunda imagem, localizada na parte inferior da figura, foi feita a mesma busca, mas com um simples erro de digitação. É possível observar que o site não recupera nenhum tipo de material, mesmo faltando apenas uma única letra.

Os livros recuperados na busca (Figura 5) possuem uma página própria na plataforma com todas as informações técnicas (título, autor, resumo, avaliações, resenhas, títulos semelhantes e etc.), disponíveis para o leitor ter mais conhecimento a respeito do livro e poder interagir com outros usuários. Assim como encontrar outros livros com temáticas semelhantes e poder adicionar em seu acervo pessoal (Figura 6).



Figura 6. Página do livro “O mundo assombrado pelos demônios”

Na Figura 6 é possível observar parte da página do livro dentro do Skoob. Nela são apresentadas as avaliações feitas pelos usuários, os comentários e resenhas, livros com assuntos semelhantes, sinopses e abaixo das sinopses os gêneros literários que são considerados como termos representativos dos assuntos: “Filosofia”, “não-ficção” e “Sociologia”, mas esses termos não possibilitam serem acessados para que redirecionem para outros livros com as mesmas temáticas.

A análise dos critérios e procedimentos da indexação e da busca e recuperação de livros no Skoob possibilitou perceber a existência das abordagens cognitiva e baseada no usuário mencionadas por Brandt e Medeiros (2010), para os quais o sistema de representação surge a partir dos dados criados pelos usuários e, ao mesmo tempo, se relaciona com as operações da recuperação da informação, que também são desempenhadas pelos usuários, refletindo diretamente o universo cognitivo do usuário.

Enquanto que no *GoodReads*, apesar de haver um bibliotecário que realiza a indexação, não existe nenhuma informação a respeito de como é feito esse procedimento ou se há uso de algum tipo de vocabulário. Os usuários têm permissão apenas de atribuírem termos representativos dentro da sua estante e, a partir delas, acessar ao acervo pessoal dos outros usuários, mas não no registro geral.

## 5. Conclusões

A partir da pesquisa bibliográfica e da análise dos resultados na plataforma do *GoodReads*, foi possível observar que a indexação é feita apenas pelo bibliotecário, mas não existe informações de como é feito e nem se há uso de vocabulário controlado. O *GoodReads* não permite que os usuários realizem a indexação social e nem acessem os metadados dos registros gerais, mas possibilita que os usuários atribuam termos representativos dentro de suas estantes pessoais, no qual podem fazer trocas e acesso de informações com outros usuários, com o objetivo de encontrar interesses literários em comuns. Foi possível notar falta de controle de vocabulário, devido a repetição de termos apresentados nas listas, muito semelhante à lista de termos autorizados.

Já no Skoob quem realiza a indexação social são apenas usuários (leitores, autores e editoras). Os termos representativos são constituídos por gêneros literários e *tags*, sem a existência de nenhum tipo de controle de vocabulário. Mas é importante ressaltar que, no Skoob e no *GoodReads*, é fornecida uma lista muito simples de gêneros literários para ser selecionado no momento do cadastramento do livro. Nessa lista não é possível fazer nenhuma inclusão de termos e não há nenhuma informação a respeito de sua origem e de sua alimentação.

O levantamento da literatura e análise dos resultados permitiu observar que a temática indexação social é uma área recente, em constante desenvolvimento, abrindo oportunidade para mais investigações, pois há problemas relacionados a terminologia, causado pela variação terminoló-

gica e conceitual, em que diferentes autores definem o mesmo tema com termos diferentes. É importante acrescentar que os ambientes de indexação social são, também, variados e, em muitos casos, desconhecidos pelos profissionais da informação, que poderiam participar com contribuições relevantes e adequadas. Dessa forma, sugere-se que os estudos acerca desse assunto continuem sendo pesquisados, contribuindo para o desenvolvimento da temática na área.

## Referências

- Amaral, Adriana; Salvador, Tarciso. (2018). Folksonomia em Sites de Redes Sociais Segmentadas (SRSS) em livros: um estudo exploratório da interface do GoodReads. // Rev. Digital de Biblioteconomia e Ciência da Informação (RBDCl). 16:2 (maio/ago. 2018) 397-413.
- Assis, Juliana de; Moura, Maria Aparecida (2013). Folksonomia: a linguagem das tags. // Encontros Bibli: revista eletrônica de Biblioteconomia e Ciência da Informação. 18:36 (2013) 85-106.
- Barros, L. M. S. (2011). A folksonomia como prática de classificação colaborativa para a recuperação da informação. 2011. Dissertação (Mestrado em Ciência da Informação) – Universidade Federal do Rio de Janeiro: Instituto Brasileiro de Informação em Ciência de tecnologia, Rio de Janeiro, 2011.
- Brandt, Mariana; Medeiros, Marisa Brascher Basílio (2010). Folksonomia: esquema de representação do conhecimento?. // Transinformação. 22: 2 (2010) 111-121.
- Burlamaque, Fabiane Verardi; Barth, Pedro Afonso (2015). Redes sociais e o ensino: o skoob como ferramenta para o letramento digital e literário. // Nuances: Estudos sobre Educação. 26:3 (2015) 53-73.
- Chang, Hsia-Ching; Iyer, Hemalata (2012). Trends in twitter hashtag applications: design features for value-added dimensions to future library catalogues. // Library trends. 61:1 (2012) 248-258.
- Dziekaniak, Gisele; Pacheco, Roberto; Kern, Vinicius Medina (2011). Revisitando a organização do conhecimento através dos tesouros, folksonomias e ontologias: aportes da engenharia do conhecimento. // Silva, Fabiano Couto Corrêa da; Sales, Rodrigo de (Orgs.). Cenários da organização do conhecimento: linguagens documentárias em cena. Brasília: Thesaurus, 2011, 201-237.
- E-Commerce Brasil (2020). Internet durante a pandemia: 97% dos entrevistados a usam todos os dias. E-Commerce Brasil. (2020).
- Fiorin, José Luiz (2006). Introdução ao pensamento de Bakhtin. São Paulo: Ática, 2006.
- Gómez-Díaz, R. (2012). Etiquetar en la web social. Barcelona: Editorial UOC, 2012.
- Gonçalves, J. L. C. S; Assis, J. de (2016). A indexação social enquanto prática de representação colaborativa da informação imagética: a construção da memória na plataforma Flickr. // Revista Conhecimento em ação. 1:2 (jul./dez 2016).
- González-Fernandez-Villavicencio, N. (2019). Reading indicators on the social networks Goodreads and LibraryThing and their impact on Amazon.// Zeitschrift für Katalanistik, 32 (2019). 143-167.
- Guedes, Roger de Miranda (2010). A abordagem dialógica na indexação social. Dissertação de Mestrado, Universidade Federal de Minas Gerais, Belo Horizonte, 2010.

- Guedes, Roger de Miranda; Moura, Maria Aparecida; Dias, Eduardo José Wense (2011). Indexação social e pensamento dialógico: reflexões teóricas. // *Informação & Informação*. 16:3 (jan./jun 2011). 40-59.
- Hassan-Montero, Yusef (2006). Indización social y recuperación de información. // *No solo usabilidad Journal*, 5. (novembro 2006).
- Holstrom, Chris (2018). Social taggins: organic and retroactive folksonomies. // *Joint Conference on digital libraries*, 18. 2018. 179-182
- IAB – Brasil (2014). Hábitos de consumo de mídia em 2013. // IAB – Brasil. 2014.
- IAB – Brasil (2020). Relatório institucional. // IAB – Brasil. 2020.
- Kipp, M. E. I.; Beak, J.; Graf, A. M. (2015). Tagging of banned and challenged books. // *Knowlegde Organization*. 42:5 (2015). 276-283.
- Kroski, E. (2005). The Hive mind: folksonomies and user-based tagging. 2005.
- Recuero, R. (2011). *Redes sociais na internet*. 2. ed. Porto Alegre: Sulina, 2011.
- Santos, Raimunda Fernanda dos; Corrêa, Renato Fernandes (2015). Modelos colaborativos de indexação social e sua aplicabilidade em bibliotecas digitais. // *Liinc em Revista*, 11:1 (maio 2015). 273-286.
- Shirky, C. (2005). *Ontology is overrated: categories, links and tags*. 2005.
- Skoob (2021). Quem Somos. Skoob. [https://www.skoob.com.br/inicio/quem\\_somos](https://www.skoob.com.br/inicio/quem_somos)
- Sousa, R. J. P. L.; Benetti, M. (2016). A indexação social no Skoob: a descrição do livro ‘O mundo de Sofia’ a partir de tags e resenhas. // *Informação & Informação* 21:1 (2016). 520-541.
- Sterling, B. (2005). What’s the best way to tag, bag, and sort data? // *Give in to the unorganized masses*. 2005.
- Vaidya, P.; Harinarayana, N. S. (2016). The role of social tags in web resource Discovery: an evaluation of user-generated keywords. // *Annals of Library and Information Studies* 63 (2016). 289-297.
- Vander Wal, Thomas (2015). *Explaining and showing broad and narrow folksonomies*. 2015.

---

Enviado: 2021-03-30. Segunda versão: 2021-06-09.  
Aceptado: 2021-06-17.

---



# La organización del conocimiento al servicio de la interconexión entre las instituciones de la memoria y el turismo cultural

*Knowledge organization for the interoperability among memory institutions and cultural tourism*

Francisco Javier GARCÍA MARCO

Universidad de Zaragoza, Instituto de Patrimonio y Humanidades, Pedro Cerbuna 12, 50009 Zaragoza (España), jgarcia@unizar.es

## Resumen

Se plantean las oportunidades para una alianza estratégica entre las industrias del turismo, la memoria cultural y la edición en el marco del despliegue de la web semántica y se aborda el papel central de la organización del conocimiento en este proyecto. Tras un análisis de la situación de cada una de estas industrias y de sus sinergias, se plantea la aportación de la organización del conocimiento en dos direcciones: mediante la aplicación del análisis de dominio a la modelización del sistema de colaboración, y del saber acumulado en la construcción de estructuras de conocimiento no redundantes y en el mapeo de sistemas conceptuales para la interconexión de sistemas, con especial atención a las posibilidades que ofrece la norma ISO 25964. El análisis se ejemplifica con un estudio de caso sobre la ecología informacional web del Camino de Santiago, y se presentan las lecciones aprendidas en el proceso.

**Palabras clave:** Organización del conocimiento. Turismo. Industria editorial. Instituciones de la memoria. GLAM. Web semántica. Interoperabilidad semántica. ODS. Camino de Santiago.

## 1. Las nuevas necesidades del sector turístico: una oportunidad para las instituciones de la memoria y la organización del conocimiento

Aunque la organización del conocimiento es por su propia naturaleza transdisciplinar —tanto por los aportes que recibe como por las áreas de aplicación—, es cierto que sus centros y áreas de práctica más formalizados y consolidados están especialmente ligados al campo de las instituciones de la memoria: bibliotecas, archivos, museos y sitios de referencia. Sin embargo, la explosión de la web, la proliferación de esquemas de organización del conocimiento fuera de la tradición de la disciplina (taxonomías, menús, anillos de sinónimos, ontologías, etc.) y la creciente interacción de especialistas de muy diversas disciplinas en torno al problema común de la organización del conocimiento —con sus diferentes puntos de vista y terminologías—, hacen necesario atender

## Abstract

The opportunities for a strategic alliance among the tourism, cultural memory and publishing industries within the framework of the semantic web are discussed; and the central role in this project of knowledge organization as a discipline is addressed. After an analysis of the situation of each one of these industries and their potential synergies, the contribution of knowledge organization is proposed in two directions: through the application of domain analysis to the modelling of the whole collaboration system; and of its know-how in the construction of non-redundant knowledge structures and in the mapping of conceptual systems to system interoperability, with special attention to the possibilities offered by the ISO 25964 standard. The analysis is exemplified with the case of the web informational ecology of the Saint James' Way, and the lessons learned in the process are presented.

**Keywords:** Knowledge Organization. Tourism. Publishing industry. Memory institutions. GLAM. Semantic web. Semantic interoperability. ODS. Saint James' Way.

al nuevo universo de información digital e integrar viejos y nuevos saberes y prácticas.

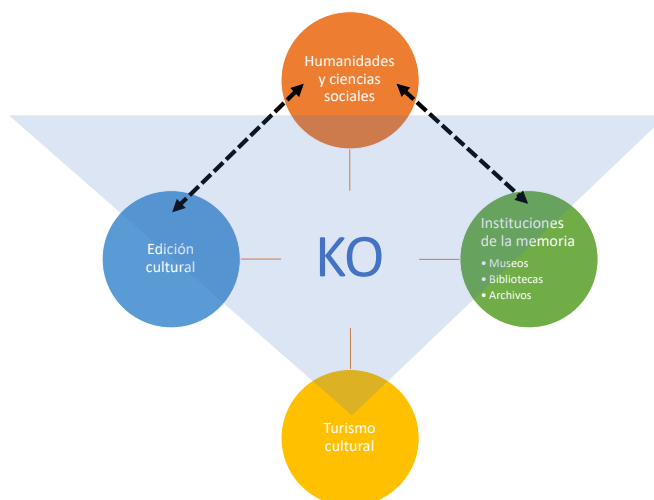


Figura 1. El potencial de la interoperabilidad a explorar

En este contexto parece especialmente interesante la gran oportunidad que ofrece el acercamiento en el nuevo contexto digital entre una de las principales industrias de nuestra época —el turismo— y las instituciones de la memoria al servicio de unos turistas cada vez más educados y en busca de experiencias más personales, contextualizadas y genuinas.

### 1.1. La importancia de la industria turística y su giro cultural

Es difícil exagerar la importancia actual del turismo como industria (Lowry, 2017, p. xxxii). En 2019, antes de la crisis Covid-19, el World Travel & Tourism Council (2021a) cifraba su contribución al PIB mundial en el 10,4%, aportando nada más y nada menos que una décima parte de los trabajos del mundo. En España, país eminentemente turístico, la aportación del PIB dicho año ascendía al 14,1% y la contribución al empleo al 14,4%, esto es, casi uno de cada siete empleos (World Travel & Tourism Council, 2021b).

El sector se encontraba en franco crecimiento, pues en marzo de 2015, cuatro años antes, las cifras se situaban un poco por debajo del 10% del PIB y de 1 de cada 11 puestos de trabajo (Lowry, 2017, p. Xxxii, siguiendo al WTTC). De hecho, entre 2014 y 2019 el turismo aportó uno de cada cuatro trabajos nuevos del mundo (World Travel & Tourism Council, 2021a).

Aunque la pandemia le ha asestado un duro golpe bajando su contribución al 5,5% del PIB mundial y volviendo a las cifras de empleo de 2015 (1 de cada 11 trabajos), se espera una rápida recuperación en cuando mejore la situación sanitaria, para convertirse en “the key sector in driving the recovery of the global economy post COVID-19” (World Travel & Tourism Council, 2020).

Pero el turismo no solo es una cuestión de interés económico, sino también de intercambio humano y realización personal. La actividad turística de tantos millones de personas y la rápida incorporación de los pueblos al turismo conforme alcanzan el suficiente nivel de ingresos (Fang, 2020) muestran la decisiva valoración que le otorgan tanto las personas como sus organizaciones. Además, es difícil negar su papel en el conocimiento de la cultura y las condiciones de vida de otras personas en un contexto de relación positiva, contribuyendo así de forma decisiva a través de la superación de los prejuicios y al establecimiento de interacciones positivas, la fraternidad y la paz, como, más allá de la indudable utilización propagandística del tópico y de la complejidad de los diferentes contextos, muestra la experiencia europea (Bechman, 2020, p. 402).

Esto es especialmente cierto en el caso del turismo cultural, que la World Tourism Organization (2012) define a grandes rasgos como el movimiento de personas por motivaciones culturales, como los viajes de estudio, los viajes para asistir a eventos culturales, lugares y monumentos, para estudiar la naturaleza, el folklore o el arte, y las peregrinaciones, incluyendo tanto el patrimonio material como inmaterial, lo que incluye aspectos como la música, la gastronomía, la lengua y otros aspectos de la vida cotidiana o de la cultura de élite. Es difícil desde esta definición que el turismo cultural no quede incorporado a casi cualquier viaje turístico en mayor o menor manera —no en vano *sightseeing* se traduce en muchos países hispanos como ‘turistear’—, por lo que los académicos se han concentrado sobre todo en estudiar los tipos de turistas culturales, que varían en un continuo que va del ocasional o anecdótico al explorador o experto dedicado, normalmente en función de su motivación (McKercher and Du Cross, 2002; Petroman, 2013). En cualquier caso, el impacto del turismo en la valoración y preservación del patrimonio cultural tangible e intangible es, a pesar de constituir una relación compleja con sus claroscuros, es decisiva; y ese efecto se extiende por tanto a las instituciones y personas implicadas en esta tarea.

Ahora bien, bajo el impacto de la revolución en las tecnologías de la información —fundamentalmente, de la World Wide Web y las plataformas móviles— se ha ido produciendo un cambio muy importante en la naturaleza del turismo. En 2015 Xiang y sus colaboradores (2015) señalaban que se estaba produciendo “a growing ‘bifurcation’ between traditional online travellers, i.e., those who use the Internet for standard travel products, and those who are beginning to adopt alternative channels and products in search of deeper and more authentic experiences”. En términos de la matriz del Boston Consulting Group, aunque el mercado tradicional seguía constituyendo la “vaca”, el ligado a las nuevas posibilidades que ofrece la Internet para localizar productos y experiencias más personalizados se estaba convirtiendo en la “estrella”. Este viraje hacia un turismo más inquisitivo, educado y personalizado en busca de experiencias más profundas ligadas a la personalidad del turista ha llevado a expertos del campo a diagnosticar un ‘cultural turn’ en la industria (Dabbage 2018a, 55-56; 2018b).

El impacto radical que ha tenido la epidemia Covid en la actividad turística de masas se ha unido al creciente consenso internacional para una actividad económica sostenible concretado en la Agenda 2030 (Naciones Unidas, 2015) para enfatizar la necesidad de contar con alternativas al turismo de masas, basado en productos estándar



igualmente masivos. En España y en otros países está contribuyendo también a este giro cultural del turismo la creciente preocupación por la despoblación de las áreas fundacionales de las actuales sociedades posmodernas, frecuentemente referido en nuestro país con el eslogan de la “España vacía”. Finalmente, una población más diversa en inquietudes y fundamentos busca también cada vez más una experiencia turística más segmentada y auténtica desde su personal punto de vista.

## 1.2. Las humanidades, las instituciones de la memoria y la edición: un sector en redefinición

Durante la larga historia de la civilización, el esfuerzo de preservación del conocimiento y la evolución de las tecnologías de la información documental han ido resultando en diferentes especialidades profesionales, que en diferentes épocas de la historia también se han ido agrupando como consecuencia de las urgencias coetáneas. Así las bibliotecas se especializaron en época helenística desde los archivos, pero con las desamortizaciones de la época moderna surgieron cuerpos de funcionarios conjuntos que abordaban el conjunto del patrimonio desamortizado; igualmente, con el desarrollo tecnológico en época contemporánea las bibliotecas explotaron en multitud de “tecas” y archivos especializados, pero poco a poco las “tecas” se han ido reagrupando también bajo sus especialidades matriz — bibliotecas y archivos— en forma de departamentos, colecciones y profesionales especializados. Con la explosión de la World Wide Web ha pasado algo semejante, y hoy en día las emergentes bibliotecas digitales vuelven a verse como un tipo de biblioteca especializado o parte de una biblioteca híbrida en vez de como una alternativa que cuestiona la trayectoria anterior o se separa de ella.

En el momento actual, desde un punto de vista meramente sincrónico, en los países avanzados, las leyes de depósito legal y las iniciativas de archivo privadas han hecho en las últimas décadas que todo el sector de la producción cultural —edición de libros y revistas, documentos de archivo, producción audiovisual, videojuegos, sitios web, etc.— se fueran convirtiendo de iure en industrias culturales, pues a través de ellas se ha estipulado la preservación de sus contenidos y su inclusión en las instituciones de la memoria.

Desde un punto de vista diacrónico, ha sido la revolución digital la que ha acercado todo este sector que de nuevo se vuelve a ver como una unidad, especialmente desde el punto de los informáticos de sistemas de gestión del patrimonio y

de los gestores de más alto nivel que se ven obligados a tomar una perspectiva amplia, ya sea por transversalidad de las aplicaciones y sistemas, ya por la jerarquización lógica de la política cultural. Como la digitalización se centra en la creación de representaciones documentales de todo el universo del patrimonio digital, contribuye a documentar la totalidad del ámbito de la preservación y difusión del patrimonio, que se aprecia como una unidad por parte de gestores, informáticos y profesionales de la información. Por fin, la perspectiva web además obliga a integrar los recursos hacia el usuario, que en general los ve como ‘recursos digitales’ al mismo nivel, lo que ha contribuido a la integración y compatibilización de esquemas de metadatos —si quiere, reglas de catalogación— y vocabularios de búsqueda — sistemas de organización del conocimiento (SOC)— (Given and McTavish, 2010). Todos estos factores han favorecido que un término tan artificial como GLAM (Galleries, Libraries and Archives and Museums) haya ido ganando momento en los últimos años. (Esta tendencia se puede apreciar realizando una simple búsqueda retrospectiva con los términos TOPIC:(GLAM galleries libraries archives museums) —en plural para buscar artículos genéricos, filtrando otros sentidos de la palabra ‘glam’, notablemente la abreviatura de glamorous— en Web of Science (WoS), por ejemplo.

En el campo de la Information Science, algunos autores han escogido el término de instituciones de la memoria (Hjerppe, 1994; Hjørland, 2000) para referirse a ese amplio mundo de instituciones que se ocupan de la preservación del conocimiento, y efectivamente a nivel social a asegurar la memoria colectiva, y que se tomará también aquí como referencia (García Marco, 2010). También este término ha ido ganando momento en la última década, como se puede apreciar realizando la búsqueda TOPIC:(“memory institutions”) (en plural para buscar artículos genéricos sobre el tema) en WoS.

Ciertamente, desde un punto de vista más gerencial, se viene refiriendo a todas estas instituciones bajo el paraguas de la preservación del patrimonio, como instituciones patrimoniales, fijándose más en su dimensión material heredable entre las generaciones. Ese mundo también está sufriendo una gran transformación por el impacto de la globalización y las tecnologías de la información, y la desaparición de los modos de vida tradicionales. De hecho, la Unesco ha asumido con energía la incorporación del patrimonio inmaterial —esto es, todavía no codificado o almacenado de forma organizada— dentro de su campo de acción. El impacto de las nuevas tecnologías

ha producido el despegue de un campo innovador al que se suele referir con el concepto de 'patrimonio digital' o *'digital heritage'*.



Figura 2. Un sector GLAM ampliado

Por otra parte, en este artículo se defiende que el dominio de trabajo de las instituciones de la memoria debe coordinarse estrechamente con los productores y transformadores de la información que aquellas custodian por varias causas. La primera razón es que la revolución digital está haciendo que mucha parte del trabajo de formateo y etiquetado de los datos y documentos se vaya desarrollando cada vez más dentro de las propias empresas editoras y de los propios centros de investigación, y todo hace pensar que este proceso aumente con la expansión de la web semántica y el movimiento de las publicaciones y datos abiertos enlazados. En segundo lugar, las empresas editoras y los centros de investigación —especialmente, pero no solo, los relacionados con las humanidades— son a la vez proveedores y consumidores privilegiados de las instituciones de la memoria, siendo incluso el origen de muchas ellas —por ejemplo, los archivos y bibliotecas académicas—, por lo que son sus aliados naturales en el proceso de cambio tecnológico en el que están inmersas. Por eso, en este artículo se propone el uso ampliado del acrónimo GLAM++, para denotar la necesidad e incorporar a estos dos grupos de instituciones al esfuerzo de normalización e interoperabilidad. Para este sector ampliado, la denominación de instituciones de la memoria parece especialmente adecuado, pues trasciende la función patrimonial para abordar la investigación y el descubrimiento (que, por otra parte, muchas instituciones preservadoras también realizan). Por esta razón, los usaremos de forma indistinta con este significado a partir de ahora.

Por otra parte, la necesidad de competir en una economía global está haciendo que los fondos dedicados a la investigación se centren cada vez más en objetivos de rentabilidad económica, lo que pone en peligro disciplinas enteras que acumulan siglos de reflexión y desarrollo en el campo de las humanidades, y someten el mundo de las ciencias sociales a una extraordinaria presión de las modas del momento —forzadas por la dinámica política— que dificultan la acumulación orgánica del saber.

Para asegurar la supervivencia de estas disciplinas en estas condiciones, resulta necesario conectarlas con fuentes de financiación fuertes y estables compatibles con el nuevo énfasis en el desarrollo económico. En esta dirección, las industrias culturales y especialmente el turismo por su potencia económica son apuestas seguras, como muestran respectivamente el éxito de las series históricas y de los grandes museos en las últimas décadas. Ciertamente, será necesaria mantener la distinción entre ciencia aplicada y pura para moderar la instrumentalización de estas disciplinas (que, por otra parte, siempre se ha producido), pero al menos será más fácil mostrar su relevancia económica y social y asegurar su sustento económico.

### 1.3. Oportunidades para una alianza estratégica entre el sector de la memoria cultural y del turismo

Se ha discutido en las páginas anteriores como tanto el mundo del turismo como el de las humanidades y la memoria cultural están sometidos a presiones sociales, económicas y tecnológicas que ponen en cuestión la eficacia de su gestión actual. La globalización pone en peligro con su poder de aculturación la pervivencia del patrimonio cultural, pero a la vez despierta en las masas cada vez mejor educadas un hambre por el conocimiento de sus raíces lejanas y el encuentro con otras culturas y formas de vida. Las tecnologías de la información actúan como un enorme nivelador al difundir los consensos alcanzados por los poderosos a naciones y pueblos en rápida transformación, pero a la vez permiten la gestión, preservación y difusión de enormes cantidades de información, y su descubrimiento de forma altamente personalizada.

La idea es que, a medida que la información patrimonial y el conocimiento humanístico se exprese con los nuevos formalismos que permiten su procesamiento automático con fines diversos —según los principios y estándares del movimiento de datos abiertos enlazados—, toda esa información se ponga al servicio (y también sirva

de contraste y contexto, por qué no) de la producción cultural y de la explotación turística del patrimonio cultural material e inmaterial. Por su parte, el sector del turismo podrá beneficiarse de información más contextualizada, personalizable, contrastable y genuina. Por otro lado, el sector GLAM++ podrá “mejorar su visibilidad y relevancia a través de una contribución incontestable, aplicada y práctica al desarrollo económico y social” (García-Marco, Galindo, Lasala y López-del-Ramo, 2020), pensando especialmente en la necesidad de afrontar la rápida decadencia y despoblación de los espacios rurales y de las ciudades pequeñas que han constituido la base de nuestras culturas, y que no es sino manifestación de una ola imparable que va transformando el mundo entero desde que empezara a manifestarse en la Inglaterra de la industrialización (Laslett, 1965).

Las fuentes y el saber humanístico “datificados” pueden alimentar el torrente de datos que un turismo enriquecido, preparado y acompañado por páginas web inteligentes y aplicaciones de realidad aumentada y virtual necesita para adaptarse

a turistas cada vez más preparados, que buscan una experiencia personalizada y enriquecida, dándoles acceso al nivel de complejidad que necesitan y para el que están preparados. Gracias a este aporte de información a través de dispositivos móviles, los espacios físicos se convierten en entornos accesibles, comunicables, didácticos, recomendables, contrastables y comparables, permitiendo al visitante sumergirse en las lecciones ocultas que encierran, tanto en el momento de seleccionar los destinos, como en la preparación del viaje, la propia visita, y la documentación y comunicación de la experiencia.

La arquitectura básica de este esfuerzo se sintetiza en la Figura 3: se trata de movilizar las fuentes culturales y de infraestructuras turísticas teniendo en cuenta la segmentación de los usuarios para proporcionarles un mensaje ajustado a sus necesidades a través de diversas aplicaciones y plataformas de comunicación con el objeto de que puedan alcanzar una experiencia enriquecida en las diferentes fases de la actividad turística.



Figura 3. Una visión ganar-ganar para turistas, destinos, agentes turísticos e instituciones de la memoria

En conclusión, las humanidades y las instituciones de la memoria se pueden beneficiar recíprocamente de la llegada de recursos para hacer posible este esfuerzo y, en general, facilitar la financiación de la investigación básica al conectarla mejor con sus usos económicos potenciales. Esto parece especialmente importante en una época en que el acelerado cambio cultural, las luchas político-culturales entre identitarios y globalistas y la sucesión de crisis económicas (2008, 2020) han contribuido a poner en peligro la inmediata relevancia y financiación de la cultura frente a las urgencias económicas y sanitarias del momento a lo largo de un periodo muy prolongado.

#### 1.4. La organización del conocimiento al servicio de la integración de las industrias culturales

La organización del conocimiento puede contribuir de una forma sustancial a la visión de unas instituciones de la memoria y servicios turísticos conectados de forma más estrecha y productiva. A nuestro entender, puede hacerlo fundamentalmente de dos maneras. En primer lugar, contribuyendo a hacer inteligible el esfuerzo a través de una de sus metodologías más propias, el análisis de dominio (Hjørland & Albrechtsen, 1995; Hjørland, 2017), aplicado sobre todo sobre el conjunto de sitios web en los que se expresa el quehacer de las instituciones turísticas y de la memoria, entendidos como ecología informacional (García Marco, 2015). En segundo lugar, contribuyendo a conectar entre sí a los diferentes

agentes gracias a la sólida comprensión de los principios universales y compartidos de la clasificación que la organización del conocimiento como disciplina aporta (Ranganathan, 1937; Dahlberg, 1995) y de los instrumentos para compatibilizar estructuras conceptuales para hacerlas interoperables (Zeng, 2019). Esfuerzo que debe abordarse actualmente en el marco de la web semántica en cuanto que infraestructura capaz de conectar datos de forma a la vez procesable automáticamente por las máquinas e inteligible por los humanos, apoyándolo desde la capa de la web semántica que le es más propia a la organización del conocimiento: la ontológica. Las dos siguientes secciones del artículo están específicamente dedicadas a abordar sendas vías.

## 2. La modelización de la ecología informacional del turismo cultural: agentes y organizaciones del conocimiento en concurrencia

El concepto de ecología aplicado al mundo de la información y la comunicación puede ser extraordinariamente fructífero, y tiene una tradición de varios decenios en el campo de la ciencia de la información (Hubermann 2001; Shim and Lee 2006; Sebastián 2008). Igual que sucede en el ámbito de los ecosistemas biológicos, los diferentes sistemas de información que coexisten en un sistema social establecen relaciones de coexistencia, cooperación y competencia entorno a las necesidades de información presentes en él, formando un sistema complejo. Esto se aprecia especialmente en el mundo de la Internet, en el que resulta relativamente fácil estudiar con métodos informétricos (*webmetrics*) su evolución. En definitiva, una ecología informacional se puede definir como un conjunto de organismos que interaccionan entre ellos y con su ambiente para resolver necesidades de información, y cuyo análisis cuidadoso deber ser siempre el paso previo de cualquier intervención futura. Estas ecologías informacionales que evolucionan de una forma propia contrastan con los sistemas de información diseñados, con las arquitecturas informacionales; pero, por supuesto, también pueden ser intervenidas y normalizadas hasta cierto punto, como sucede por ejemplo con otros sistemas abiertos como los sistemas eléctricos o de comunicaciones.

Un buen ejemplo es la información sobre el Camino de Santiago en la web, en la que se puede apreciar con nitidez ese conjunto de elementos en interacción dinámica en función de un objetivo (comunicarse con los usuarios interesados en la ruta, desde diferentes puntos de vista) y también sus disfunciones (López del Ramo and García Marco, 2018). En la figura 4 se pueden apreciar sus componentes fundamentales agrupados por

sectores, cada uno haciendo énfasis en los servicios que más les interesa ofrecer.



Figura 4. Agentes de la ecología informacional del Camino de Santiago

Desde el punto de vista de los objetivos de este artículo, se trataría de incorporar a esta ecología surgida de una manera natural el input de las instituciones de la memoria en función de los intereses de los turistas (usuarios). Quedarían por tanto tres subsistemas a analizar —entrada desde las instituciones de la memoria, mediadores de información turística, y turistas— que se van a analizar en detalle en las tres siguientes subsecciones desde el punto de vista de su interoperabilidad semántica.

### 2.1. Los creadores de información: las instituciones de la memoria

Quizá el mayor reto de movilizar la información de las instituciones de la memoria hacia el sector turístico sea la gran complejidad y diversidad de los esquemas de metadatos —en parte paliado por la posibilidad de utilizar esquemas de “máximo común divisor” como Dublin Core— y, centrándonos en lo que nos ocupa en este artículo, los SOC utilizados. Por eso, en este trabajo la cuestión se ha dividido en dos fases: en primer lugar, conocer los SOC utilizados en cada tipo de instituciones de la memoria; y, en segundo, lugar explorar estrategias para su interoperabilidad.

Las bibliotecas han sido durante décadas y siguen siendo el sector más avanzado en la estandarización, aunque archivos y museos están acercándose rápidamente bajo la presión y el incentivo de la digitalización y la publicación de sus fondos en Internet. Las bibliotecas comparten un

esquema de metadatos muy elaborado y normalizado (la familia MARC21 y las DRA), y en menor medida un conjunto de SOC en gran parte complementarios (sistemáticos y alfabéticos), pero también en competición: LCC, DC, CDU, entre otros, por la parte de los SOC sistemáticos; y LCSH, RAMEAU, EMBNE, entre otros, por la parte de los sistemas de encabezamientos de materia, y algo semejante sucede con las autoridades de persona, entidad y lugar. Sin embargo, programas de mapeo entre todos ellos están en curso desde hace años y en muchos casos muy avanzados (Slavic, 2011; Zen, 2019).

Otros problemas surgen cuando se intentan utilizar para conectar los recursos con las necesidades de los mediadores turísticos y los propios turistas. Mientras que los encabezamientos de materia suelen ser suficientemente específicos, las clasificaciones no lo son tanto por su alcance muy general y fundamentalmente disciplinar. Los encabezamientos de lugares, entidades y personas resultan ser con mucho los más específicos.

Esto se aprecia fácilmente con el ejemplo del Camino de Santiago (García, Galindo, Lasala y López, 2020). Por lo que respecta a las clasificaciones, si nos centramos en la más usada en su ámbito geográfico, se observa el Camino no dispone una clase en la CDU: normalmente se clasifica en la clase de Geografía Regional (913), y en menos ocasiones en la de Rutas (656.022) o en las muchas y dispersas clases disponibles para Viaje (por ejemplo, 338.48-12, 656.022.33), ya que transporte y turismo se encuentran en clases diferentes. Para darle especificidad a unas clasificaciones tan generales, en las bibliotecas españolas se suele añadir una faceta de país (44+460, Francia y España, por lo de Camino Francés, y aunque el Camino no esté limitado a estos países). Los bibliotecarios de la Biblioteca Nacional de España haciéndose cargo del problema, han venido añadiendo a los recursos relevantes la cadena “Camino de Santiago” entre corchetes después de los códigos de país, pero como se observa en los catálogos colectivos del país esta es una solución particular de esta institución. Por el contrario, por lo que respecta a los sistemas de encabezamientos de materia, aunque tampoco existe un encabezamiento específico, al menos son fácilmente reconocibles y, si fuera necesario, mapeables:

- EMBNE: Peregrinaciones cristianas — Santiago de Compostela
- LCSH: (Christian pilgrims and pilgrimages-Spain-Santiago de Compostela)  
RAMEAU: [Pèlerinages chrétiens-Espagne-Saint-Jacques-de-Compostelle (Espagne)]

El otro problema que plantean los recursos bibliotecarios es que ofrecen documentos completos —información sistematizada y organizada como un todo—, lo cual resulta muy interesante para ayudar al turista a documentarse en general o sobre ciertos aspectos concretos. Por el contrario, no permiten un acceso rápido a consultas concretas —exigen un costoso ojeo—; y, sobre todo, no permiten agregar y manipular datos, que son la materia prima de las decisiones turísticas en la Internet. El costoso proceso de convertir libros, artículos y documentos audiovisuales en datos estructurados llevará años y necesitará el concurso de los centros de investigación y de las editoriales, como hemos comentado antes.

Sin embargo, los museos —lógicamente nos referimos a los que tienen los medios necesarios— descansan para su complicada gestión en bases de datos muy analíticas, que les han servido para situarse entre las instituciones de la memoria en la vanguardia de la aplicación de las tecnologías semánticas. En el campo de los museos ha surgido también una sofisticada propuesta ontológica —CIDOC CRM—, que ha trascendido su ámbito para convertirse en una referencia fundamental en la publicación como datos enlazados de los recursos digitales de las instituciones de la memoria en su conjunto. En el ámbito museístico han surgido también SOC muy analíticos y bien organizados como el Art and Architecture Thesaurus (AAT) o el Iconclass, entre otros.

La situación del mundo archivístico respecto a la propuesta de una interconexión más estrecha entre las instituciones de la memoria y el sector turístico es más distante. Los turistas, excepto el tipo deliberado que se considerará más tarde, raramente interactúan directamente con los fondos y documentos, salvo en exhibiciones, de forma muy parecida a como lo hacen con los objetos museísticos. El uso de documentos de archivos por parte de los turistas es secundario respecto a otros activos, pues no se emplea directamente en la planificación del viaje y en las decisiones turísticas. Como es bien conocido, los usuarios de archivos son fundamentalmente investigadores y administraciones y personas con necesidades de prueba legal. Además, los archivos tienen formas de organización del conocimiento más idiosincrásicas, basadas en el principio de proveniencia y agrupación natural de los fondos, aunque se dispone también de alternativas más transversales como la clasificación funcional. Tesoros y encabezamientos tienen un papel menor, pero su uso está creciendo transversalmente en el conjunto de las instituciones de la memoria sobre todo en lo que se refiere a los sistemas de autoridades de nombres de persona, entidad y lugar. En cualquier caso, y a pesar de las razones antedichas,

la interrelación de los documentos y fondos de archivos —y de los datos que contienen— no debería faltar en la estrategia que se está planteando, especialmente pensando en los turistas más deliberados y en la producción de exhibiciones y contextualizaciones virtuales y de realidad aumentada.

Como se ha señalado, el sector de la producción editorial —impresa y audiovisual— tendrá que transformarse en los próximos años para acomodarse dentro del nuevo marco de transmisión del conocimiento al que apuntan las tecnologías semánticas y sociales, con sus frentes más visibles en la actualidad: los datos enlazados, el big data y las aplicaciones de la inteligencia artificial. Para que esto sea posible los autores y los editores tienen que proveer sus documentos en forma anotada, de manera que sus diferentes niveles de datos sean accesibles para su combinación y procesamiento automático. Las bibliotecas y sus esquemas de metadatos también se tendrán que ir transformando para facilitar y preservar este nuevo mundo de información. La tendencia actual a la convergencia de medios y reutilización de contenidos favorecen estos procesos de marcado profundo dentro de los proyectos editoriales.

## 2.2. Los mediadores de información: DMOs, asociaciones, empresas y particulares

En el caso de la información turística, y aunque en ocasiones las instituciones de la memoria se comunican directa y eficazmente con los turistas —notablemente los museos y centros de interpretación—, su circulación se produce fundamentalmente a través de los mediadores turísticos, conocidos técnicamente como organizaciones de marketing o comercialización de destinos (en inglés, destination marketing organizations, DMOs) —esto es, las oficinas de promoción turística de los estados, regiones, y especialmente comarcas y municipios—, y cada vez más a través de las agencias de viaje en línea y sus redes sociales específicas (Xiang et al., 2015), que compiten cada vez con más fuerza con las agencias tradicionales, que también se han ido volcando en el nuevo medio. Además ciertas empresas, especialmente editoriales especializadas en turismo y patrimonio funcionan también en la práctica como DMOs y servicios de documentación y comunicación de los DMOs oficiales.

Por fin, las asociaciones de expertos y aficionados en cuestiones culturales, así como ciertas personas individuales también juegan un papel importante en la información turística cultural. En el caso de nuestro ejemplo son notablemente importantes las asociaciones de amigos del Camino de Santiago, y autores que mantienen blogs de

mucho interés. La internet ha permitido la democratización de la publicación, y actualmente pequeñas organizaciones y personas muy dedicadas pueden tener una presencia global notable.

La organización del conocimiento para los mediadores de información turística tiene dos dimensiones: una interna, relacionada con la gestión del conocimiento de la organización (o del individuo) en sus diferentes aspectos —captura o representación del conocimiento en información, preservación, recuperación, y transformación de la información en conocimiento— a través de un SAC corporativo o personal; y otra externa, dirigida a comunicar de forma organizada la información de interés a los turistas, de manera que sean informados, puedan convertir esa información en conocimiento y finalmente tomar un conjunto de decisiones turísticas que se concreten en un viaje. La comunicación externa se hace normalmente a través de sitios web y plataformas sociales, aunque cada vez más se proponen a los usuarios también aplicaciones (app) móviles crecientemente sofisticadas. En los sitios web se utiliza para organizar el conocimiento fundamentalmente taxonomías que se manifiestan sobre todo en los menús de usuario; y en los sitios sociales sobre todo folksonomías y, en algunos casos, taxonomías ad hoc. En las aplicaciones web y móviles se detecta también el uso de ontologías subyacentes en los casos más avanzados, sobre todo en los sistemas de recomendación (Abrahams & Dai, 2005; Li, Buhalis & Zhang, 2013; Grün, Neidhardt & Werthner, 2017).



Figura 5. Las dos dimensiones de la organización del conocimiento en los mediadores turísticos

Aunque ambas dimensiones —que coinciden con la tradicional distinción entre intranet, internet

(y extranet)— tienen que estar conectadas para evitar redundancias e inconsistencias, cada una debe ser abordada en sus propios términos. Los contextos internos son mucho más especializados conceptualmente, idiosincrásicos terminológicamente y se organizan fundamentalmente a partir de las funciones y la estructura organizacional. Sin embargo, los medios externos tienen que ser capaces de comunicarse eficazmente con el usuario, acercándose a su(s) esquema(s) natural(es) de organizar el dominio y a sus actividades típicas. Sin embargo, y a pesar de su lógica proximidad, muchas veces se aprecian disfunciones en ambos sentidos: o ambas dimensiones están construidas por separado, sin comunicarse y estructurarse adecuadamente; o, por el contrario, una absorbe a la otra, anulándola y frustrando sus objetivos (Figura 6, García, Galindo, Lasala y López, 2020). En particular, es muy frecuente que la taxonomía corporativa se convierta en la base del sitio web dedicado a los turistas (López del Ramo and García Marco, 2018) o que se centre básicamente en los productos ofrecidos; aunque también se encuentran ejemplos de sitios web muy bien diseñados

desde una perspectiva centrada en el usuario (Tabla I).



Figura 6. Enfoque centrado en el usuario frente a centrados en la organización en el diseño de webs turísticas

Dirección	<i>caminodeantiago.gal</i>	<i>turismocastillayleon.com</i>	<i>caminodeantiago.org</i>	<i>Santiagoways.com</i>
Descripción, sector	Sitio orientado al turista del Gobierno de Galicia	Sitio del Gobierno de Castilla y León	Sitio de la Federación Española de Asociaciones Jacobeas	Agencia de viajes
Conocer, motivar	Descubre			St James' Way
Preparación del viaje (equipo)	Prepárate	Consejos A pie En bicicleta A caballo	Consejos	Consejos A pie En bicicleta Grupos
Ruta	Planifica tu viaje	Equipamiento 'Credencial' 'Turismo España' GPS	Camino	Caminos [cada uno]
En Camino	Durante el Camino	'Correos' Información sanitaria Reciclado	Servicios	'Nuestros' servicios
Cultura	Santiago y Galicia Conocimiento e investigación		Scriptorium Links	Historia
Promotor, marca	Xunta de Galicia	Contacto	Noticias de la Federación	Firma, Nuestro equipo, Contacto, Noticias

Tabla I. Categorías principales de cuatro sitios de diferente tipo sobre el Camino de Santiago

### 2.3. Los consumidores de información: los turistas

El mundo de los turistas —los consumidores finales de la información— no es menos complejo que el de los productores y mediadores de infor-

mación (Figura 7). Aunque las necesidades de información generales y específicas son muy semejantes —planificar y contextualizar el viaje, resolver problemas prácticos, buscar datos específicos y navegar las fuentes—, no todos los turistas las persiguen con igual interés ni eficacia, ni en general ni en contextos concretos.



Figura 7. Modelización del turista y el turismo cultural: turistas, necesidades y productos

Esto es así porque, como en otras actividades humanas, los sujetos están muy segmentados en cuanto a intereses, capacidades y dedicación efectiva. Esto se aprecia tanto en las categorizaciones que se han propuesto de los turistas en cuanto a los aspectos culturales, como a los productos de marketing que se les ofrecen y que en sí mismos son identificativos de una segmentación comercial de los turistas. Así, en cuanto a los turistas y para el caso estricto del turismo cultural MacKercher y Du Cross (2002) han identificado cinco tipos diferentes: deliberados (*purposeful*), turísticos (centrados en las atracciones), afortunados (*serendipitous*), casuales e incidentales; mientras que Csapó (2012) fue capaz de identificar siete tipos diferentes de productos de turismo cultural diseñados para distintos tipos de turistas: turismo patrimonial; rutas temáticas culturales; turismo cultural urbano, excursiones culturales; turismo étnico y tradicional; turismo de eventos y festivales; turismo religioso y rutas de peregrinaje; y, finalmente, cultura creativa y turismo creativo.

Esta diversidad se refleja en la variadísima terminología utilizada por los usuarios en las búsquedas y en su desconexión (Xiang et al. 2009) con las propuestas de organización del dominio que realizan los sitios web turísticos, más allá de las típicas categorías generales del campo (acceso, alojamiento, gastronomía, sitios que ver, contexto histórico, etc.). Como señalan Xiang y sus

colaboradores (2009) “el dominio completo es extremadamente rico y en gran parte idiosincrásico [...] con muchos detalles dependientes del destino”. Afortunadamente, el lenguaje de los turistas puede ser explorado en profundidad a partir de los contenidos que ellos crean —blogs, posts y reseñas (*reviews*) (Gretzel, Hwang and Fesenmaier 2006)—; pero, por el contrario, todo este mundo es difícil de abarcar, modelizar y presentar organizadamente a los usuarios. Todo ello hace pensar que, con el tiempo, el trabajo de organizar un dominio tan complejo solo se puede ir resolviendo parcela a parcela en concurrencia por diferentes agentes, y que los abordajes generales servirán para integrar este complejo mundo pero que un resultado global sólo puede ser fruto del trabajo de muchos años.

Por otra parte, la superación de esta brecha entre los usuarios y la información turística a través de la modelización conceptual de ámbitos específicos ha sido objeto de un gran interés por especialistas en inteligencia artificial durante las dos últimas décadas y se ha convertido en un campo consolidado de investigación bajo la denominación de sistemas de recomendación turística (Grün, Neidhardt and Werthner, 2017). Estos sistemas se orientan sobre todo a establecer correspondencias de destinos y recursos turísticos con perfiles ‘ontológicos’ de los turistas a partir de características de personalidad y preferencias de viaje. La conexión de estos sistemas con fuentes



documentales y bibliográficas requerirá un trabajo muy detallado y cuidadoso, especialmente tras superar los intereses básicos de los turistas para alcanzar el difícil territorio del aprendizaje, el descubrimiento y la inmersión cultural; pero constituye un campo muy prometedor de investigación.

### 3. La ruta hacia la interoperabilidad semántica

Teniendo en cuenta el análisis anterior, se considera que, para conseguir una integración de los recursos culturales con los intereses de los turistas en el marco de las tecnologías semánticas, es necesario avanzar en al menos tres frentes: a) avanzar en la *modelización de los intereses de los usuarios* de manera que se abarque de forma a la vez comprensiva e inteligible su diversidad; b) identificar las *tecnologías disponibles* más adecuadas para la implementación del *punto ontológico*; y c) proponer *metodologías* que permitan abordar el problema por partes y sectores, pero con una perspectiva global, auténticamente *integradora*, y que sea *factible*, sin sucumbir ante la evidente complejidad de la tarea. A continuación se analizan estos tres objetivos con más detalle.

#### 3.1. La modelización del usuario y sus mediaciones

Como hemos visto en la sección 3.3 las perspectivas de los turistas y de las organizaciones turísticas sobre su actividad son extraordinariamente diversas y muy sujetas a las especificidades tanto del turista como del destino. Más allá del inventario imprescindible de los servicios turísticos —desplazamiento, alojamiento, manutención, actividades de ocio, contexto, servicios básicos (salud, seguridad, servicios sociales, etc.)— y su segmentación económica (ofertas, última hora, etc.), es necesario un modelo del turista que permita también incorporar las dimensiones experienciales y de autorrealización que una aproximación cultural al turismo requiere. Quizá para esta misión resulte especialmente útil la pirámide de las necesidades humanas propuesta por Maslow (1954), que permite incluir tanto las más básicas —desplazamiento, alojamiento, comida y bebida, seguridad—, como las sociales (identidad, relación) y las de autorrealización (creación de sentido, cultura popular y alta cultura, sabiduría y espiritualidad).

Lógicamente, una ontología de las necesidades humanas carece de la suficiente especificidad para representar un dominio concreto como el turismo salvo en sus niveles más generales, y fundamentalmente como una estructura contrastiva: esto es, de comparación y evaluación. Para las especificidades del campo, contamos con al menos

cuatro fuentes: a) el bruto de los textos, imágenes, vídeos y comentarios de las páginas web y redes sociales, que pueden ser analizados como corpora y proporcionar terminología e iconología; b) las numerosas taxonomías desarrolladas para los sitios web turísticos, ya sea estrictamente —como taxonomías de CMS— o de forma subyacente y apreciable fundamentalmente en los menús de usuario, que proporcionan también estructuras jerárquicas y auténticos mapas de conceptos; c) las ontologías que se están realizando, algunas específicamente dirigidas a modelar las necesidades de los usuarios (Li, Buhalis and Zhang, 2013), que aportan relaciones complejas; y c) algunos lenguajes documentales muy específicos como el tesoro multilingüe de la Organización Mundial de Turismo (World Tourism Organization, 2001), que pueden proporcionar un marco común, ya sea de forma contrastiva o normativa.

#### 3.2. La modelización semántica y la vía ISO 25964

El despliegue normativo de la web semántica está disponible para la mayoría de sus capas (Berners-Lee, Hendler & Lassila, 2001) desde hace una década, y los avances se suceden de forma constante, especialmente en las aplicaciones de datos abiertos enlazados y grafos semánticos (Hitzler, 2021). En particular para el tema que nos ocupa se pueden expresar sistemas de organización del conocimiento en la web semántica mediante RDF, OWL, y específicamente SKOS, una ontología definida como OWL-Full para representar la estructura básica y el contenido de esquemas conceptuales como listas encabezamientos de materia, taxonomías, esquemas de clasificación, tesauros y cualquier tipo de vocabulario controlado (Pastor, Martínez y Rodríguez, 2012; Martínez González y Alvite Díez, 2019).

Por otra parte, el desarrollo de sistemas de recomendación basados en tecnologías semánticas y específicamente ontologías se ha convertido en uno de los campos interdisciplinares más activos entre turismo y ciencia de la información. Ya en su revisión entrada la década anterior Li, Buhalis y Zhang (2013) indicaban que los “Recommender systems based on semantic web and ontology technologies are an effective method and tool to improve the quality of internet service through personalization and customization.”

Por la parte de los sistemas de organización del conocimiento, la novedad de la última década ha sido la publicación y consolidación de la nueva norma de tesauros ISO 25964 (Aitchison and Dextre Clarke, 2004; Dextre Clarke, 2012). Frente a otros SOC, los tesauros constituyen la plataforma más general para la interoperabilidad,

pues comparten con la Internet su naturaleza distribuida y autónoma dentro de unos parámetros normativos claros y operativos de uso e interacción. Destacan por su compromiso entre simplicidad y potencia, y por su gran flexibilidad (García Marco, 2016a): pues son capaces de expresar todas las relaciones fundamentales normativamente creando un marco básico de interoperabilidad mientras que dejan abierta la posibilidad de aplicar modelos más específicos y aún personalizados de relaciones. Finalmente, incorporan el multilingüismo, imprescindible en el entorno web, de forma nativa.

Además de asumir los conceptos como unidades básicas —lo que resuelve de una manera más natural la cuestión multilingüe— y de actualizar y simplificar las anteriores normas en una (ISO, 2011), la parte segunda de ISO 25964 (ISO, 2013) aborda específicamente la interoperabilidad de los tesauros con el resto de SOC, incluyendo taxonomías y ontologías, analizando los diferentes modelos de mapeo y proponiendo herramientas de mapeo específicas (=EQ, ~EQ, +, |, Bm, NM, RM). En principio, los enfoques de hub parecen más adecuados para la generación de mapeos exhaustivos de cara a portales o servicios específicos, aunque los mapeos abiertos están funcionando muy bien para crear grafos abiertos de mucho interés.

La arquitecta del esfuerzo de renovación de los tesauros, Dextre Clarke (2011) realizó muy pronto un diagnóstico preciso de los resultados esperados de los mapeos con diferentes SOC, que es extremadamente relevante para el proyecto presentado en este artículo. Por su parte, Soergel (2010) propuso un modelo conceptual pensando en las búsquedas facetadas que tan populares se han vuelto, por ejemplo, en los sitios de comercio electrónico, también de mucho interés. Como hemos visto en el apartado 3.1 los mapeos que pueden ser más productivos en el caso de la interconexión de sitios de información turística e instituciones de la memoria son, además de los propios tesauros, las listas de encabezamiento de personas, entidades, lugares, títulos y materias por su especificidad respecto a los artefactos culturales y su contexto concreto. Los niveles superiores van a ser siempre difíciles y complejos, pues su estructuración depende de la manera de abordar el dominio por cada agente, lo que no quiere decir que se deban abandonar por imposibles; antes bien, una línea de trabajo necesaria es ir explorando mapeos y proponiendo macrolenguajes de interconexión.

### 3.3. Vías de acción

La interoperabilidad es un problema complicado incluso cuando se intenta entre SOC bien establecidos, pero se vuelve especialmente difícil cuando también se intentan integrar taxonomías web y folksonomías procedentes de los sitios web de los proveedores de información turística y de las redes sociales. Por ello, es necesario encontrar vías prácticas para reducir el espacio del problema y hacerlo abordable. En sendos proyectos pilotos que tenemos en marcha—uno sobre el Camino de Santiago en Aragón y el otro sobre la ruta de Goya en Aragón— se han estado trabajando por el momento en dos direcciones, siempre con el enfoque directo de aprovechar la interoperabilidad semántica para construir y enriquecer sitios web turísticos más funcionales.

En primer lugar, se han estudiado los diferentes tipos de sitios web identificados en la sección 3 —gobiernos regionales y municipales, asociaciones de peregrinos, empresas y particulares— con el objetivo de alcanzar un “mínimo común múltiplo” de los conceptos presentes en ellos, mapeando sus relaciones jerárquicas naturales, tal y cómo aparecen en dichos sitios web. De esta forma, los conceptos quedan inventariados en su totalidad pero sin redundancia. A partir de aquí se han controlado los términos equivalentes para permitir un acceso terminológico completo a los usuarios, pues como señalan Xiang, Gretzel y Fesenmaier (2009) es imprescindible recoger el complejo lenguaje del turismo para dar servicios a tipos de turistas muy diferentes. Después se han reducido las relaciones polijerárquicas a una relación monojerárquica preferida y las relaciones asociativas necesarias para representar las alternativas, de forma que se puedan derivar taxonomías. A esta herramienta se están añadiendo mapeos con SOC especialmente relevantes, que toman el carácter de términos equivalentes, específicos o genéricos, y simples o compuestos, según el tipo de mapeo. El resultado es un tesoro multilingüe monojerárquico embebido, compatible con la producción de una taxonomía web simple que pueda servir para el diseño de un sitio web tipo. A este tesoro lo hemos denominado Tesoro Común Taxonómico Compacto (TCTC) (Figura 8 y Figura 9). En el caso de crear un sitio modelo común cada concepto tendría por un lado, un correlato en un nodo o subnodo del CMS subyacente; y una o varias sentencias de búsqueda SPARQL para referir a cada fuente de datos culturales. En este último caso, las sentencias se podrían generar de forma común partiendo del vocabulario codificado en el TCTC y las características comunes de las fuentes externas de datos, con lo que sería

mucho más fácil extender los mapeos. La capacidad que ha mostrado WikiData para convertirse en un hub semántico también de vocabularios estructurados permitiría minar sus mapeos para ampliar las fuentes de forma automática (García Marco, 2016b). Recíprocamente, de cara a facilitar el servicio de datos enlazados hacia fuera —

esto es, no limitarse sólo a utilizar datos externos, sino también servirlos—, bastaría añadir un punto SPARQL al CMS. Las especificidades de este tipo de proyectos están tratadas de forma excelente en otro artículo de este mismo número monográfico (Hernández Carrascal y Agenjo Bullón, 2021).

NivelClave	Clase	Término controlado local	Término simplificado	Término ISO	Observaciones	Categorías—Número	Término controlado corrido	
1	7C35EC50	0	[Instrumentales]			0	[Instrumentales]	
2	7637DEE0	0.4	Busquedas		Buscar	4	0.4 Busquedas	
2	7AF4EEC	0.5	Idiomas			6	0.5 Idiomas	
1	A46F0EB5	1	Camino de Santiago	El Camino	Camino de Santiago	En general. El Camino de Santiago, El Camino	13	1 Camino de Santiago
2	25872139	1.0	Caminos de Santiago (otros)			Página dedicada a los otros caminos de Santiago, salvo los que cuida esa	1	1.0 Caminos de Santiago (otros)
2	9F6855A7	1.1	Caminos de Santiago en España	Caminos	Caminos de Santiago	Enfoque general que resalta su diversidad y alternativas. Sin: Caminos	8	1.1 Caminos de Santiago en España
3	C987FC8F	1.1.0	Criterios de valoración				1	1.1.0 Criterios de valoración
3	3AF4DD0	1.1.1	Camino Francés	Camino Francés	Camino Francés		4	1.1.1 Camino Francés
5	CC38E4C	1.1.1.0.1	Etapas del Camino Francés	Etapas			1	1.1.1.0.1 Etapas del Camino Francés
5	9550C67C	1.1.1.0.2	Mapas del Camino Francés	Etapas			1	1.1.1.0.2 Mapas del Camino Francés
4	5E215F26	1.1.1.1	Camino Francés en Aragón		Ruta de Somport		3	1.1.1.1 Camino Francés en Aragón
5	6B3947AF	1.1.1.1.1	Camino Francés en Aragón (Puerto de Palo)		Ruta del Puerto de Palo		1	1.1.1.1.1 Camino Francés en Aragón (Puerto de Palo)
4	CEFD32A	1.1.1.2	Camino de Santiago en Estella	Estella	Camino de Santiago		1	1.1.1.2 Camino de Santiago en Estella
5	E11038FF	1.1.1.2.1	Etapas del Camino de Santiago en Estella				7	1.1.1.2.1 Etapas del Camino de Santiago en Estella
5	05502B9F	1.1.1.2.2	Puntos de interés en Estella				1	1.1.1.2.2 Puntos de interés en Estella
4	970A2813	1.1.1.3	Camino de Santiago por Burgos	Burgos (provincia)	Camino de Santiago		1	1.1.1.3 Camino de Santiago por Burgos
5	B898DB1	1.1.1.3.1	Camino de Santiago por Burgos (ciudad)	Burgos (ciudad)	Camino de Santiago		1	1.1.1.3.1 Camino de Santiago por Burgos (ciudad)
6	D1F26710	1.1.1.3.1.1	Iconografía jacobea en Burgos	Burgos (ciudad)	Camino de Santiago		1	1.1.1.3.1.1 Iconografía jacobea en Burgos (ciudad)
5	52A34AEA	1.1.1.4.1	Etapas del Camino de Santiago en el Bierzo				2	1.1.1.4.1 Etapas del Camino de Santiago en el Bierzo
4	4E64813C	1.1.1.5	Camino de Santiago en Palencia				1	1.1.1.5 Camino de Santiago en Palencia
5	9F538FD	1.1.1.5.1	Etapas del Camino de Santiago en Palencia				1	1.1.1.5.1 Etapas del Camino de Santiago en Palencia
4	0B6B7CA8	1.1.1.6	Camino de Santiago por León	León			2	1.1.1.6 Camino de Santiago por León
4	A454BDC	1.1.1.7	Camino de Santiago del Bierzo	El Bierzo	Camino de Santiago		1	1.1.1.7 Camino de Santiago del Bierzo
4	C4AC5D8	1.1.1.8	Camino de Invierno				2	1.1.1.8 Camino de Invierno
5	BE757CE3	1.1.1.8.1	Etapas del Camino de Invierno				1	1.1.1.8.1 Etapas del Camino de Invierno
5	7ABEF104	1.1.1.8.2	Mapas del Camino de Invierno				1	1.1.1.8.2 Mapas del Camino de Invierno
4	D5FCA066	1.1.1.9	Camino de Fisterra-Muxía				2	1.1.1.9 Camino de Fisterra-Muxía
5	E2788D6B	1.1.1.9.1	Etapas del Camino de Fisterra-Muxía				1	1.1.1.9.1 Etapas del Camino de Fisterra-Muxía
5	6CA46B0B	1.1.1.9.2	Mapas del Camino de Fisterra-Muxía				1	1.1.1.9.2 Mapas del Camino de Fisterra-Muxía
3	09D892A0	1.1.2	Caminos de Santiago del Norte	Caminos del Norte	Caminos de Santiago		1	1.1.2 Caminos de Santiago del Norte
4	D24AC667	1.1.2.1	Camino del Norte	Camino del Norte	Camino del Norte		4	1.1.2.1 Camino del Norte

Figura 8. Vista de ejemplo de la base de datos con la taxonomía común construida para las Asociaciones de Amigos del Camino de Santiago

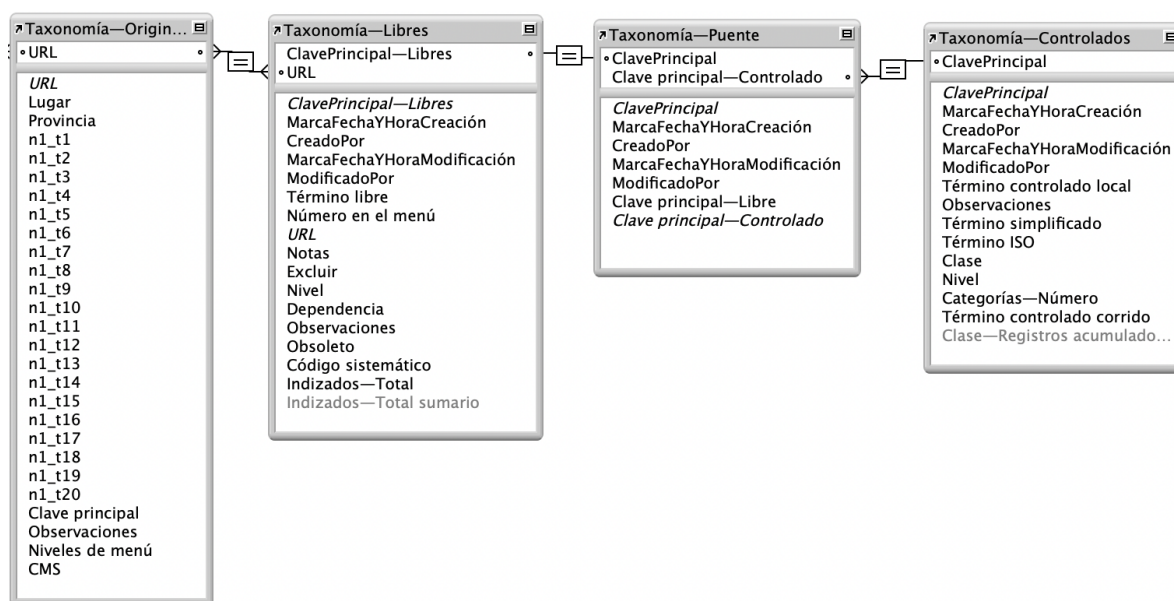


Figura 9. Estructura de la base de datos para controlar los taxones y reexpresarlos

En una fase más avanzada se pretende explorar estructuras facetadas y recuperar las polijerarquías para obtener menús adaptados para los diferentes tipos de usuarios. Por el momento se ha aplazado este abordaje, en seguida se vio que en una estructura totalmente facetada la terminología se alejaba mucho de la usada cotidianamente por los usuarios, pues la descoordinación de conceptos complejos para facetarlos genera mucha abstracción y ambigüedad desde el punto de vista del que desea realizar búsquedas específicas. Sí sería posible en el futuro diseñar un sistema facetado que funcionara de forma subyacente para expresar y comunicar diferentes alternativas de organización del dominio turístico en cuestión, como se ha señalado al principio del párrafo.

Por otro lado, se han ido explorando los diferentes KOS existentes en el campo de las instituciones de la memoria para comprobar su potencial para la interoperabilidad con sitios web turísticos. Como hemos visto, una vez mapeada la correspondencia de los diferentes KOS con los nodos de TCTC es posible generar sentencias y filtros de mantenimiento automático. Mapear los lenguajes entre sí y generar de forma general un hub común se demostró una tarea desbordante para pequeños y medianos proyectos, que, además, no resulta productiva en muchos vocabularios debido a la especificidad de la terminología necesaria para servir a los usuarios de sitios e intereses específicos. Díaz-Corona, Lacasta y Nogueras (2021) presentan en este mismo número monográfico los resultados de un mapeo automático del sitio InfoGoya con diversos vocabularios que resulta muy ilustrador al respecto. Lo dicho no invalida que un mapeo en hub que incorpore tanto los KOS de las instituciones de la memoria como de los agentes turísticos pueda constituir un servicio abierto de enorme valor, lo que probablemente se acometerá en un estado de desarrollo más avanzado del dominio que nos ocupa. De hecho el proyecto que se presenta es compatible con su evolución hacia una estructura de este tipo.

Un punto débil de toda esta estrategia es que asume a priori que los SOC de las instituciones de la memoria y de las organizaciones turísticas representan adecuadamente los intereses, conceptos y vocabulario de los usuarios. A esta objeción hay que señalar, por una parte, que la estrategia de mapeo no invalida la necesidad de seguir mejorando los SOC para ajustarlos a sus diferentes tipos de usuarios. Es importante realizar con asiduidad estudios de usuarios tanto subjetivos (satisfacción, entrevistas, etc.) como objetivos (log analysis, search analysis, eye-tracking, entre otros), y de hecho son ya una parte muy

importante de las investigaciones en organización del conocimiento. Por otra parte, la propia naturaleza de la labor de mapeo, que pone en contacto sistemas diferentes, sirve para compensar las debilidades de unos sistemas con las fortalezas de otros, ampliando y nivelando el mapa de conceptos resultante.

#### 4. Conclusiones y perspectivas futuras

En este artículo se ha intentado presentar un análisis de un campo transdisciplinar emergente de gran interés —la conexión entre industrias de la memoria, editoriales y turísticas mediante tecnologías semánticas—; así como las lecciones ganadas en un caso práctico de aplicación al prototipado de un sitio web de información turística enriquecido con datos culturales, realizando una valoración de sus posibilidades, limitaciones y algunos enfoques alternativos. Se puede concluir que mapear estructuras conceptuales de diferentes dominios, que además pueden ser abordados desde perspectivas de usuario muy diferentes, es un reto difícil, que visibiliza grandes lagunas comparativas entre dominios, y que resulta muy intensivo en trabajo intelectual y humano a pesar de las herramientas tecnológicas disponibles, especialmente si se pretende un ajuste fino. Sin embargo, constituye el futuro de la web y una promesa razonable de mejora —al menos parcial— a los graves problemas de redundancia, dispersión y manipulación de la información que se aprecian en la actualidad.

Sobre todo, se ha intentado mostrar que la interconexión de recursos culturales y necesidades de la industria turística mediante tecnologías semánticas auxiliadas desde la disciplina de la organización del conocimiento es un campo de colaboración transdisciplinar de gran interés para todas las partes implicadas: las instituciones de la memoria pueden ganar en relevancia económica y visibilidad en un entorno de investigación competitivo y muchas veces refractario; la industria turística puede ganar en posibilidades de personalización a una demanda turística cada vez más sofisticada; y las tecnologías semánticas y la organización del conocimiento campos de prueba de gran interés, llenos de retos relevantes y generalizables.

Pero además, este esfuerzo de ajuste fino entre los diferentes sectores implicados constituiría un aporte significativo a la sostenibilidad natural y cultural, ayudando a conectar de forma inteligente demanda y oferta ahorrando recursos, y a preservar la riqueza natural y cultural a través de su comunicación sistemática, pero dirigida, a los usuarios.

En definitiva, este campo de investigación emergente puede también constituir una gran aportación de la organización del conocimiento como disciplina a un turismo más intensivo y menos extensivo, más respetuoso y comprometido con la sostenibilidad de la naturaleza y las personas, lo que se ha denominado turismo responsable; y demostrar el gran potencial que la organización del conocimiento tiene para aportar metodologías que favorezcan un desarrollo más sostenible en el marco del consenso generado en torno a la Agenda 2030 (Naciones Unidas, 2015).

## Agradecimientos

Este artículo publica resultados de investigación financiados con el proyecto CSO2015-65448-R (MINECO/FEDER). Se agradece sinceramente el apoyo a las instituciones financiadoras sin cuya generosidad esta investigación no hubiera sido posible.

## Referencias

- Abrahams, Brooke and Wei Dai (2005). Architecture for Automated Annotation and Ontology Based Querying of Semantic Web Resources. // 2005 IEEE/WIC/ACM International Conference on Web Intelligence, Proceedings, 413-417.
- Aitchison, Jean and Stella Dextre Clarke (2004). The Thesaurus: A Historical Viewpoint, with a Look to the Future. *Cataloging & Classification Quarterly*. 37, 5-21.
- Bechmann Pedersen, S. (2020). A Passport to Peace? Modern Tourism and Internationalist Idealism. // *European Review*. 28:3, 389-402. doi:10.1017/S1062798719000516
- Berners-Lee, T.; Hendler, J.; O. Lassila (2001). The semantic web. // *Scientific American*. 284:5, 76-88.
- Csapó, Janos (2012). The Role and Importance of Cultural Tourism in Modern Tourism Industry. // *Strategies for Tourism Industry: Micro and Macro Perspectives*, edited by Murat Kasimoglu and Handan Aydin. London: IntechOpen. 201-232.
- Dahlberg, Ingetraut (1995). Current trends in knowledge organization. // *Organización del conocimiento en sistemas de información y documentación : actas del I Encuentro de ISKO-España, Madrid, 4 y 5 de noviembre de 1993* (pp. 7-26): Zaragoza : [ISKO-España : Universidad de Zaragoza]. [http://www.iskoiberico.org/wp-content/uploads/2014/07/007-026\\_Dahlberg.pdf](http://www.iskoiberico.org/wp-content/uploads/2014/07/007-026_Dahlberg.pdf)
- Debbage, Keith (2018<sup>a</sup>). Economic Geographies of Tourism: A Critical and Contested Discourse. // *The SAGE Handbook of Tourism Management*. London: SAGE. 53-68.
- Debbage, Keith (2018<sup>b</sup>). Economic Geographies of Tourism: One 'Turn' Leads to Another. // *Tourism Geographies*. 22, 347-353.
- Dextre Clarke, Stella G. (2011). In Pursuit of Interoperability: Can We Standardize Mapping Types?. // *Concepts in Context: Cologne Conference on Interoperability and Semantics in Knowledge Organization*; held 19-20 July 2010; Cologne, Germany, edited by F. Boteram, W. Goedert, and J. Hubrich. Würzburg: Ergon Verlag.
- Dextre Clarke, Stella G. (2012). ISO 25964: A Standard in Support of KOS Interoperability. // *Facets of Knowledge Organization*, 4-5 July 2011, London / eds. Alan Gilchrist and Judi Vernau. London: Emerald, 129-134.
- Díaz-Corona, Dayany; Lacasta, Javier; Noguera Iso, Javier (2021). Enriquecimiento automático de portales culturales mediante modelos de organización del conocimiento. // *Scire*. 27:1.
- Fang, Wei Tang (2020). *Tourism in Emerging Economies: The Way We Green, Sustainable, and Healthy*. Singapore: Springer.
- García Marco, Francisco Javier (2010). Las instituciones de la memoria ante el cambio digital: una propuesta sobre la articulación interdisciplinar de las ciencias de la documentación. // *Pliegos de Yuste: revista de cultura y pensamiento europeos*. 11-12, 61-70.
- García Marco, Francisco Javier (2015). Apropriación y uso de la información y del conocimiento en el contexto de las ecologías informacionales. // González de Gómez, María Nérida; Rabello, Rodrigo (org.). *Informação: agentes e intermediação*. Brasília: IBICT, 2015. p. 313-353. <http://livroaberto.ibict.br/handle/123456789/1068>. ISBN 978-85-7013-107-2.
- García Marco, Francisco Javier (2016a). The evolution of thesauri and the history of knowledge organization: Between the sword of mapping knowledge and the wall of keeping it simple. // *Brazilian Journal of Information Studies: Research Trends*. 10:1, 1-11. <http://www2.marilia.unesp.br/revistas/index.php/bjis/article/view/5786>.
- García Marco, Francisco Javier (2016b). Enhancing the visibility and relevance of thesauri in the web: searching for a 'hub' in the linked data environment. // *Knowledge Organization*. 43:3, 193-202.
- García-Marco, Francisco Javier; Galindo, Fernando; Lasala, Pilar; López del Ramo, Joaquín (2020). Advancing the Interoperability of the GLAM+ and Cultural Tourism Sectors through KOS. // *Advances in Knowledge Organization*. 17, 151-160. doi.org/10.5771/9783956507762-151. [https://www.ergon-verlag.de/isko\\_ko/downloads/aiko\\_vol\\_17\\_2020\\_18.pdf](https://www.ergon-verlag.de/isko_ko/downloads/aiko_vol_17_2020_18.pdf)
- Given, L., M. And McTavish, L. (January 2010). What's Old Is New Again: The Reconvergence of Libraries, Archives, and Museums in the Digital Age. // *The Library Quarterly*. 80:1, 7-32. <http://www.jstor.org/stable/10.1086/648461>.
- Gretzel, Ulrike, Yeong-Heyon Hwang, and Daniel R. Fesenmaier (2006). A Behavioural Framework for Destination Recommendation Systems Design. // *Destination Recommendation Systems: Behavioural Foundations and Applications*, edited by D. R. Fesenmaier, K. Wöber and H. Werthner. Wallingford, UK: CABI, 53-64.
- Grün, Christoph, Julia Neidhardt, and Hannes Werthner (2017). Ontology-Based Match-making to Provide Personalized Recommendations for Tourists. // *Information and Communication Technologies in Tourism 2017*, edited by R. Schegg and B. Stangl B. Cham.: Springer, 3-16.
- Hernández Carrascal, Francisca; Agenjo Bullón, Xavier (2021). El turismo cultural y la digitalización del patrimonio en un entorno Linked Open Data: el caso de Europa. // *Scire*. 27:1.
- Hitzler, Pascal (2021). A review of the semantic web field. // *Communications of the ACM*. 64:2, 76-83. <https://doi.org/10.1145/3397512>
- Hjerppe, R. A. (1994). A framework for the description of generalized documents. // *Advances in Knowledge Organization*. 4, 173-180.
- Hjørland, B. (2000). Documents, memory institutions and information science. // *Journal of Documentation*. 56:1, 27-41. DOI: 10.1108/EUM0000000007107.
- Hjørland, Birger (2017). Domain analysis. // *Knowledge Organization*. 44: 6, 436-464. [https://www.isko.org/cyclo/domain\\_analysis](https://www.isko.org/cyclo/domain_analysis)
- Hjørland, Birger; Albrechtsen, Hanne (1995). Toward a new horizon in information science: Domain analysis. // *Journal of the American Society for Information Science*. 46: 6, 400-425.

- Huberman, B. A. (2001). *The laws of the web: patterns in the ecology of Information*. Cambridge, Mass.: MIT Press.
- ISO (2011). ISO 25964-1:2011. Information and documentation. Thesauri and interoperability with other vocabularies. Part 1: Thesauri for information retrieval. Geneva: ISO.
- ISO (2013). ISO 25964-2:2013. Information and documentation. Thesauri and interoperability with other vocabularies. Part 2: Interoperability with other vocabularies. Geneva: ISO.
- Laslett, Peter (1965). *The world we have lost: England before the industrial age*. London: Methuen.
- Li, Nao; Buhalis, Dimitrios; Zhang, Lingyun (2013). "nterdisciplinary Research on Information Science and Tourism. // Information and Communication Technologies in Tourism 2013, edited by L. Cantoni and Z. Xiang. Berlin, Heidelberg: Springer, 302-313.
- López del Ramo, Joaquín; García Marco, Francisco Javier (2018). El Camino de Santiago en los sitios web de las Comunidades Autónomas: análisis del Contenido, orientación y encuadres temáticos predominantes. // *Revista General de Información y Documentación* 28, 703-26.
- Lowry, Linda L. (2017). Introduction. // Lowry, Linda L. (ed.). *The SAGE international encyclopedia of travel and tourism*. Thousand Oaks, Ca.: SAGE, 2017. Xxxiii-xliii.
- Martínez González, María Mercedes; Alvite Díez, María Luisa (2019). Thesauri and Semantic Web: Discussion of the evolution of thesauri toward their integration with the Semantic Web. *IEEE Access*. 7. DOI: 10.1109/ACCESS.2019.2948028
- Maslow, Abraham H. (1954). *Motivation and personality*. New York: Harper.
- McKercher, B.; Du Cross, Hilary C. (2002). *Cultural Tourism: The Partnership Between Tourism and Cultural Heritage Management*. New York: Hayworth Hospitality Press.
- Naciones Unidas (2015). *Transforming our world: the 2030 Agenda for Sustainable Development: Resolution adopted by the General Assembly on 25 September 2015 (A/RES/70/1)*. New York: Naciones Unidas. <https://en.wikipedia.org/wiki/File:N1529189.pdf>.
- Pandur, Ioana (2013). Types of Cultural Tourism. // *Scientific Papers: Animal Science and Biotechnologies*. 46:1, 385-388.
- Pastor-Sánchez, Juan Antonio; Martínez-Mendez, Francisco-Javier; Rodríguez, J. V. (2012). Aplicación de SKOS para la interoperabilidad de vocabularios controlados en el entorno de linked open data. // *El Profesional de la Información*. 21:3,245-253. DOI: 10.3145/epi.2012.may.04
- Petroman I.; Petroman, Cornelia; Marin, Diana; Ciolac, Ramona; Văduva, Loredana;
- Ranganathan, S. R. (1937). *Prolegomena to library classification*. Madras, London: The Madras library association; E. Goldston, Ltd.
- Sebastiá, Montserrat (2008). La Ecología de la Información: Un Nuevo Paradigma de la Infoesfera. // *Pliegos de Yuste*. 7-8: 24-36.
- Shim, Seonyoung and Byungtae Lee (2006). Evolution of Portals and Stability of Information Ecology on the Web. // *Proceedings of the 8th International Conference On Electronic Commerce*. New York: ACM, 584-588.
- Slavic, Aida (2011). Classification Revisited: A Web of Knowledge. // *Innovations in Information Retrieval: Perspectives for Theory and Practice*, Eds. Allen Foster and Pauline Rafferty. London: Facet. 23-48
- World Tourism Organization (2001). *Thesaurus on Tourism & Leisure Activities* (Trilingual: English, French, Spanish). Madrid: World Tourism Organization.
- World Tourism Organization (2012). Summary study on tourism and intangible cultural heritage. Madrid: UNWTO. <https://doi.org/10.18111/9789284414796>.
- World Travel & Tourism Council (2020). *Travel & tourism global economic impact & trends 2020*. WTTC, June 2020. <https://wtcc.org/Portals/0/Documents/Reports/2020/Global%20Economic%20Impact%20Trends%202020.pdf?ver=2021-02-25-183118-360> (2021-05-04).
- World Travel & Tourism Council (2021a). *Economic Impact Reports*. <https://wtcc.org/Research/Economic-Impact> (2021-05-04).
- World Travel & Tourism Council (2021b). *Spain 2021 Annual Research: Key Highlights*. <https://wtcc.org/Research/Economic-Impact/moduleId/704/itemId/206/controller/DownloadRequest/action/QuickDownload> (2021-05-04).
- Xiang, Zheng, Gretzel, Ulrike; Fesenmaier, Daniel R. (2009). Semantic Representation of Tourism on the Internet. // *Journal of Travel Research*. 47, 440-53.
- Xiang, Zheng, Dan Wang, Joseph T. O'Leary, and Daniel R. Fesenmaier (2015). Adapting to the Internet: Trends in Travelers' Use of the Web for Trip Planning. // *Journal of Travel Research*. 54, 511-27.
- Zeng, Marcia Lei (2019). Interoperability. // *Knowledge Organization*. 46:2, 122-146. <http://www.isko.org/cyclo/interoperability>.

---

Enviado: 2021-06-09.  
Aceptado: 2021-06-16.

---

---

# Redes y comunidades de descriptores en artículos de Biblioteconomía y Ciencia de la Información (1971-2020): análisis de su evolución temporal mediante Técnicas de Análisis de Redes

*Descriptor networks and communities in Library and Information Science articles (1971-2020): analysis of their temporal evolution through Network Analysis Techniques*

---

**Carlos G. FIGUEROLA, Modesto ESCOBAR MERCADO,  
Ángel ZAZO RODRÍGUEZ, José Luis ALONSO BERROCAL**

Universidad de Salamanca. Instituto de Estudios en Ciencia y Tecnología, c/Espejo 2, 37007 Salamanca, España,  
{ figue | modesto | zazo | berrocal } @usal.es

## Resumen

En este artículo se examinan las palabras clave con que los autores describen sus propios trabajos académicos, a partir de los artículos del campo temático Library and Information Science del WoS, entre 1971 y 2020. Mediante algoritmos de búsqueda de comunidades propios del análisis de redes se han identificado los principales subcampos temáticos de investigación en esta disciplina y se han analizado sus evoluciones temporales a través del uso de redes dinámicas. Los resultados muestran que algunos descriptores aparecen de manera habitual en el intervalo temporal, pero asociados con diferentes descriptores a lo largo de los años, lo cual es indicativo de una especialización temática. Existen también grupos de descriptores auto-contenidos, dibujando subcampos temáticos bien definidos desde sus primeras apariciones. Por otra parte, encontramos comunidades de descriptores bien perfiladas y con metodologías bien asentadas en unas determinadas fechas, que se van expandiendo en el tiempo, añadiendo otros descriptores de especialización, que, asimismo, permiten conectar con otras comunidades temáticas. De las trece áreas encontradas, las principales se han agrupado en torno a seis amplios descriptores interrelacionados: *bibliometrics*, *library*, *e\_government*, *e\_commerce*, *information\_literacy* y *knowledge\_management*, cuyas evoluciones se manifiestan con diferentes pautas.

**Palabras clave:** Descriptores. Palabras clave. Técnicas de análisis de redes. NetCoin. Biblioteconomía y documentación. Bibliometría. Detección de comunidades. Redes dinámicas.

## 1. Introducción

Es relativamente frecuente, en artículos y otros trabajos de tipo académico, la inclusión de palabras clave que, eventualmente, ayuden en las búsquedas a la recuperación de tales trabajos.

## Abstract

This paper examines the keywords selected by its authors to describe their own scholarly work in the subject field of Library and Information Science collected in the Web of Science portal between 1971 and 2020. The main thematic subfields of research in this discipline have been identified by means of communities search algorithms of network analysis, and their temporal evolution has been analyzed through the use of dynamic networks. Results show keywords with high frequency, but associated with another keywords evolving over time, suggesting specialized topics. There are, besides, thematic groups of keywords well defined; among the thirteen areas found, the main ones have been grouped around six broad interrelated descriptors: *bibliometrics*, *library*, *e\_government*, *e\_commerce*, *information\_literacy* and *knowledge\_management*, whose evolutions are manifested in different patterns.

**Keywords:** Descriptors. Keywords. Network analysis techniques. NetCoin. Library and information science. Bibliometrics. Community detection.

Aunque diversos editores, al igual que los mantenedores de repositorios y bases de datos bibliográficas en general pueden aplicar sus propios encabezamientos de materias, sistemas de clasificación y otros instrumentos descriptivos del contenido temático de los trabajos académicos,

utilizando para ello lenguajes controlados; las palabras clave asignadas por los propios autores de los trabajos se caracterizan por ser descriptores libres. Los autores pueden aplicar las palabras clave sin constreñirse a ningún vocabulario específico, más allá de una cierta normalización, en algunos casos, en el uso de número, mayúsculas y similares.

Aunque este tipo de palabras clave adolece de importantes problemas (Leiva, 2002) que, teniendo en cuenta la potencia de los sistemas de recuperación actuales, ha puesto en duda su utilidad para búsquedas por los usuarios, parece que pueden constituir una buena muestra de cómo los propios autores describen sus trabajos de investigación. Las bases de datos bibliográficas, cada vez con mayor y mejor cobertura, nos proporcionan una buena fuente con la que abordar el análisis sistemático de los descriptores o palabras clave de autor (Granda et al., 2003). Un análisis de los descriptores más utilizados podría proporcionar una buena imagen de la estructura temática de la investigación en un área científica determinada; y, habida cuenta de que las publicaciones tienen fecha, también nos informaría de la evolución temporal de esa estructura temática.

De otro lado, es difícil entender una determinada disciplina científica abstrayéndola de la forma en que ésta se comunica. El papel definitorio de la comunicación científica ha sido hecho notar por numerosos autores desde diversos puntos de vista (Woolgar and Latour, 1979; Leydesdorff, 2002). De esta manera, los descriptores incorporados a los propios artículos forman parte de dicha comunicación.

Sin embargo, las palabras clave no tienen valor solamente de forma individual; un descriptor por sí sólo no expresa de manera suficiente el campo temático de un artículo académico. Los descriptores se aplican en conjunción con otros, de manera que la coincidencia o uso conjunto nos proporciona un grado más de información. Varias técnicas estadísticas permiten trabajar con este fenómeno; pero las Técnicas de Análisis de Redes parecen especialmente adecuadas para llevar a cabo este tipo de estudio (Leydesdorff, 2021). Especialmente si se dispone de herramientas que faciliten un análisis diacrónico de ese uso conjunto de palabras clave; ya se ha indicado que la evolución temporal es un elemento significativo.

Así pues, el objetivo de este trabajo es analizar el uso de palabras clave del autor en los artículos académicos de un campo científico específico, el que conocemos como *Library and Information Science* (LIS); y hacerlo construyendo una red di-

námica de descriptores y analizándola con técnicas ad-hoc y con herramientas específicas capaces de manejar redes dinámicas.

Este artículo está organizado de la siguiente manera: tras la presente Introducción, se proporciona una breve revisión de otros trabajos previos sobre el mismo tema. A continuación se expondrá la metodología empleada, describiendo las fuentes de datos utilizadas y su tratamiento; al igual que los procedimientos para la construcción de la red de descriptores; también las técnicas específicas de análisis de redes aplicadas, así como las herramientas empleadas en ese análisis. Después se mostrarán los resultados obtenidos y se discutirán las implicaciones de dichos resultados. Finalmente, se ofrecerán las conclusiones obtenidas en este trabajo.

## 2. Trabajos anteriores

Los descriptores, y más específicamente, las palabras clave asignadas por los autores de artículos académicos, han sido objeto de abundante análisis desde numerosos puntos de vista. Algunos de esos análisis se han centrado en las relaciones entre las palabras clave del autor y las palabras que aparecen en título y resumen, o incluso en el texto completo. Éste es el caso de los trabajos de Gil-Leiva y Rodríguez-Muñoz (1997) o, más recientemente, los trabajos de Jaewoo y Woosan (2014) y Lu y otros (2020). Algunos analizan sus coincidencias con elementos de lenguajes controlados como los encabezamientos de materias de la Library of Congress (Strader, 2011), o con los descriptores controlados y asignados por indizadores profesionales en diversas bases de datos (Gil Leiva y Alonso Arroyo, 2005).

Algunos investigadores han estudiado la relación entre las palabras clave del autor y el éxito de los artículos; así, por ejemplo Uddin y Khan (2016) encuentran correlación significativa entre la centralidad de los descriptores en una matriz de co-ocurrencias y el número de citas recibidas. Algo parecido encontramos en Cheng y otros (2018); aquí, centrándose en los artículos publicados en la revista *Library Hi-Tech* y sus descriptores, identifican una serie de campos o temas "calientes" y documentan su evolución entre 2006 y 2017. En Lu y otros (2020), mencionado antes, también relacionan palabras clave con citas recibidas; mientras que Peset y otros (2020) estudian la supervivencia de los descriptores en función del factor de impacto de las revistas de *Library and Information Science* en Scopus entre 2005 y 2014.

Como cabe esperar, una cantidad significativa de trabajos se ocupa de detectar a través del uso de palabras clave específicas las subáreas temáticas



de disciplinas o campos de conocimiento más amplios. Para el caso de LIS, contamos con una cantidad apreciable de investigaciones; entre ellas, la ya mencionada de Cheng y otros (2018) sobre los artículos publicados en *Library Hi-Tech*, cuyo análisis de descriptores les permite afirmar que los temas predominantes en esa revista al principio del período estudiado (2005-2017), centrados en tecnología aplicada fueron evolucionando hacia aspectos más enfocados hacia usuarios y servicios prestados en las bibliotecas. Onyanha (2018) analiza un período más amplio (1971-2015) y con una base de datos más amplia: los artículos del *Web of Science* clasificados como LIS. Su análisis de las palabras clave del autor les permite distinguir diferentes subcampos de LIS, así como marcar su evolución longitudinal.

Mokhtarpour y Khasseh (2020), utilizando la misma base de datos, analizan también un período amplio, entre 1990 y 2016; mientras que en Peset y otros (2020) hacen lo propio pero a partir de la base de datos Scopus, poniendo en valor la importancia de las palabras clave del autor al carecer de las restricciones que imponen los lenguajes controlados. De otro lado, Peña (2012) trabaja sobre las palabras clave en la producción latinoamericana en archivística, mientras Rodríguez Cruz y Martínez Rodríguez (2009) lo hacen sobre el aspecto más específico de gestión de la información.

Naturalmente, encontramos análisis temáticos basados en las palabras clave del autor no sólo en LIS, sino también en otros varios campos; entre ellas, sin ánimo de ser exhaustivos, encontramos trabajos sobre sistemas de innovación (Lee y Su, 2010), investigación operativa (Duvvuru, Kamarthi y Sultornsanee, 2012), medicina (Granda y otros, 2005), tecnología educativa (Jaewoo y Woonsun, 2014), gestión e ingeniería (Chen y otros, 2016), tecnologías disruptivas (Dotsika y Watkins, 2017), ciencias sociales y humanidades en India (Tripathi y otros, 2018), turismo (Park y Jeong, 2019), "smart cities" (Min, Yoon y Furuya, 2019), recursos humanos (Yoo y otros, 2018).

El estudio de las palabras clave mediante Técnicas de Análisis de Redes Sociales (Otte y Rousseau, 2002; Van der Hulst, 2009), basándose en las coocurrencias de palabras en los mismos artículos tiene también algunos precedentes. Así, (Duvvuru, Kamarthi y Sultornsanee, 2012) aplican un modelo de red con pesos basados en el número de coocurrencias de cada par de palabras clave; el examen de esta red permite comprobar que seguía las leyes de potenciación propias de las redes libres de escala. Palabras clave con un grado elevado tendían a corresponder a descriptores transversales; mientras que era posible observar palabras clave con menor grado,

formando grupos cohesivos que se corresponderían con los diferentes campos temáticos. Jaewoo y Woonsun (2014), por su parte, modelan una red bipartita a partir de una matriz de palabras y documentos, en la que resulta posible identificar los nodos periféricos como temas emergentes. Dotsika y Watkins (2017) también construyen una red con pesos basados en la frecuencia de las coocurrencias, pero introducen en su análisis métricas como la intermediación y la proximidad, además del grado (Arif, 2015); las divergencias entre estas métricas les permiten clasificar los nodos en diferentes tipos. Lozano y otros (2019) introduce algunas métricas de detección de comunidades de palabras clave, como los *k-cores* (Doreian y Woodard, 1994) y también aplica análisis de *egonet* a palabras claves especialmente significativas.

### 3. Metodología

#### 3.1. Las fuentes de datos

Existen numerosas bases de datos bibliográficas que recogen la producción científica en los diferentes campos o disciplinas. Unas son abiertas, otras permiten el acceso a sus registros con determinadas restricciones o limitaciones; las políticas de selección e inclusión de registros, de otra parte, son también diversas. Entre estas bases de datos, una de las más conocidas es *Web of Science* (WoS); en realidad es un entramado de varias bases de datos y servicios anejos, con una trayectoria ya larga que ha pasado por diversas vicisitudes. Aunque sometida a fuertes críticas, no cabe duda de que se trata de un referente fundamental en lo que a literatura académica se refiere.

En el caso del presente estudio, se han obtenido todos los registros clasificados dentro del campo *Library and Information Science* referidos a artículos académicos sometidos a revisión por pares, desde el año 1978 hasta 2020. Esto supone un total de 114.020 artículos; sólo una parte de ellos contienen palabras clave del autor: 42.838. Las palabras clave del autor extraídas de los registros se han sometido a un proceso de normalización ligera (además de conversión a minúsculas, por un lado se han unificado manualmente aquellos descriptores que utilizaban siglas, guiones de separación o apóstrofes; por otro, se ha aplicado el lematizador de Wordnet (Miller, 1995), ya que es un lematizador poco agresivo).

El propio concepto de *Library and Information Science* puede estar sometido a sesgos, en función de la base de datos con que se trabaje. De hecho, la asignación de revistas de LIS en bases de datos como WoS o Scopus no está libre de críticas (Abrizah y otros, 2013). No es objetivo de

este trabajo entrar en esta discusión. Simplemente, se ha empleado WoS porque la mayoría de revistas tiene muy alta relación con LIS, además de que su API es sencilla de utilizar; y, en todo caso, parece que se trata de una fuente ampliamente reconocida.

Así, el conjunto total de palabras clave aplicadas por los autores es de 62.550 descriptores únicos, que suman un total de 228.420 usos. La evolución temporal del número de artículos sobre LIS, y de la parte de ellos que contienen palabras clave del autor puede verse en la Figura 1. La Figura 2 da cuenta de la evolución de la cantidad de palabras clave únicas encontradas en la colección de registros; es de destacar el formidable crecimiento de dichos descriptores únicos, mostrando una dispersión notable en la descripción temática del contenido de los artículos. Por otra parte, la media de palabras clave por artículo (teniendo en cuenta sólo los artículos que las tienen), es de 5,34 (Figura 3).

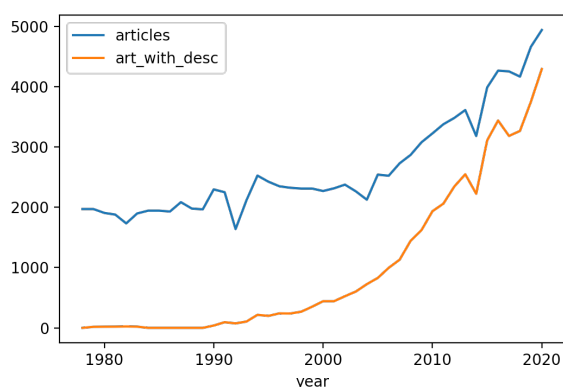


Figura 1. Evolución anual del número de artículos

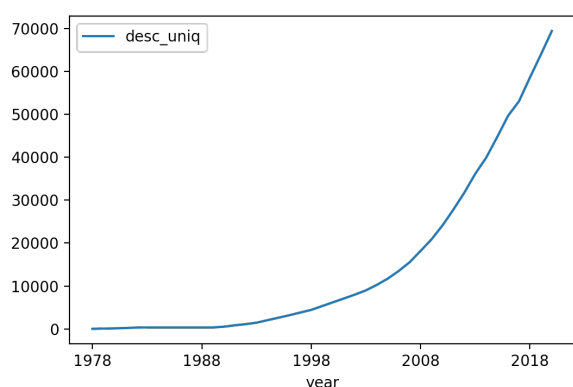


Figura 2. Evolución del número de descriptores únicos

### 3.2. Redes de descriptores

Es posible construir una red de descriptores, conectando cada uno de éstos con aquéllos con los

que es usado o coincide en un mismo artículo. La fuerza o peso de ese enlace puede considerarse proporcional a la frecuencia con que esos descriptores coinciden o son utilizados conjuntamente; de manera que la simple frecuencia de coocurrencia puede aplicarse como peso del enlace. Estos enlaces, de otro lado, carecen de dirección al ser la coocurrencia una relación bidireccional; de forma que la red resultante es un grafo no dirigido. Una red de estas características puede ser procesada mediante técnicas de análisis de redes y permite, entre otras cosas, observar el uso conjunto de varios descriptores en lugar de limitarse a la observación de palabras clave aisladas. La red resultante es relativamente grande (62.550 descriptores con 260.788 conexiones) pero puede ser podada descartando aquellas palabras clave que conectan o coinciden solamente  $n$  veces. Para  $n = 2$  obtenemos una red de 2.453 palabras clave y 7.045 enlaces, mucho más manejable.

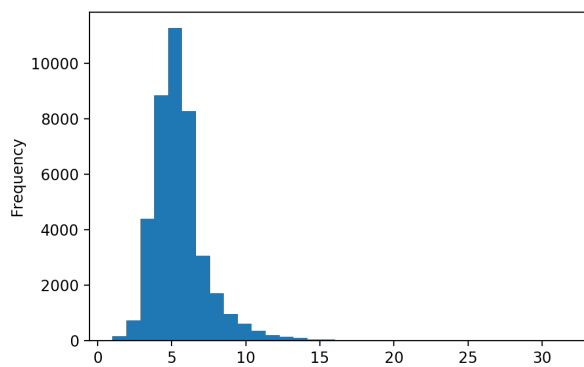


Figura 3. Número de palabras clave por artículo

La red ha sido procesada mediante la conocida librería de análisis de grafos *igraph* (versión para Python) (Csardi y Nepusz, 2006), y su análisis y representación posterior ha sido posible gracias al programa de análisis de redes *netCoin* (Escobar, 2015; Escobar y Martínez, 2020). Originariamente concebido para ser operado desde R (un conocido lenguaje de programación estadístico) puede, en realidad, ser utilizado desde otros varios lenguajes y entornos; destaca por su buena capacidad de interacción, lo que permite analizar visualmente diferentes partes y componentes de la red.

### 3.3. Red dinámica

Como se ha indicado anteriormente, la evolución temporal del uso de los descriptores es un elemento importante. Mientras la frecuencia de cada descriptor individual y su evolución temporal puede analizarse con instrumentos clásicos, la

evolución temporal de las conexiones (coincidencias) entre palabras clave y de los grupos o comunidades que esas conexiones dibujan, requieren la modelización mediante una red dinámica. Nuevamente, el software *netCoin* es capaz de representar y visualizar una red dinámica, permitiendo observar el estado de dicha red en cada momento de toda la secuencia temporal.

#### 4. Resultados y discusión

La Tabla I muestra los descriptores más frecuentes, es decir, los que aparecen en más artículos. *internet*, *social media*, *knowledge management* son algunos de los más utilizados; pero sugieren inmediatamente posibles variaciones en su frecuencia a lo largo de los años.

internet	1092
social_media	1048
knowledge_management	1034
bibliometrics	927
information_retrieval	896
academic_library	834
library	716
information_technology	571
digital_library	553
citation_analysis	552
information_literacy	535
qualitative	512
case_study	489
social_network	483
information_system	482
knowledge_sharing	481
innovation	457
china	457
trust	429
information_and_communication_technology	426
collaboration	425
electronic_commerce	412
e_government	403
research	390
public_library	368

Tabla I. Palabras clave más frecuentes

La Figura 4 muestra la evolución temporal de los más significativos; la frecuencia está puesta en relación al número de artículos recogidos cada año en la base de datos. Destaca el caso del descriptor *internet*, de elevada frecuencia de uso desde fechas muy tempranas (mediados de los años 90 del pasado siglo); su uso disminuye a partir de 2002. La palabra clave *library* sigue una trayectoria opuesta a *internet* justo en los años posteriores a la aparición en escena de éste último. Después, aunque experimenta unos años de crecimiento entra en claro declive; contrasta con el descriptor *academic\_library* que, aunque mucho menos frecuente, muestra una evolución mucho más estable. Aunque el declive de temas relacionados con las bibliotecas en general ha sido puesto ya de relieve en otros trabajos (Figueroa, García Marco y Pinto, 2017) podría pensarse que la especialización dentro del campo genérico de las bibliotecas, aunque con menor volumen, mantiene el interés de los investigadores. De alguna manera, el caso de *internet* también parece aludir a una tendencia hacia la especialización, con un decreciente contenido de la palabra clave genérica.

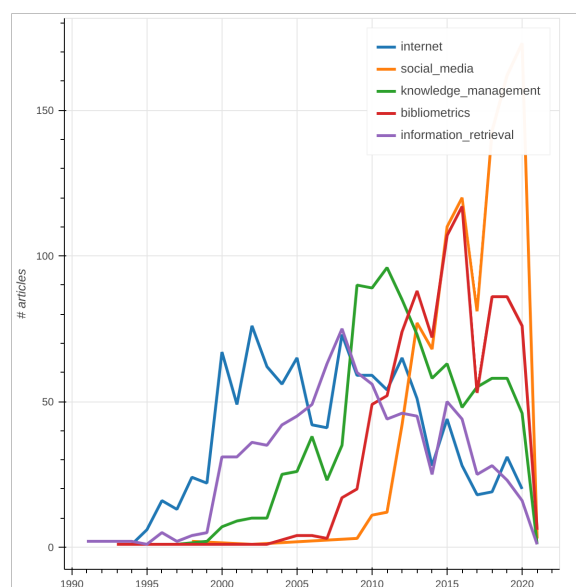


Figura 4. Evolución de palabras clave

La palabra clave *social media*, por su parte, parece un descriptor emergente, presente sólo desde 2008. En cualquier caso, puede que algunos descriptores por sí solos sean poco descriptivos, valga la redundancia, y conviene observarlos en conjunción con el resto de palabras clave con las que se aplica.

La red de todos los descriptores y sus conexiones presenta las características estructurales más relevantes que pueden verse en la Tabla II.

densidad	0.0023
diámetro	9
coeficiente medio de <i>clustering</i>	0.483

Tabla II. Métricas estructurales

De otro lado, los descriptores con mayor grado con peso (*strength*) (Arif, 2015) pueden verse en la Tabla III. Grado con peso significa, en este contexto, número de conexiones incluyendo las repetidas (dos palabras clave pueden coincidir en numerosos artículos).

Una de las aportaciones más interesantes del análisis de redes es la posibilidad de observar grupos de palabras clave que con frecuencia aparecen juntas; es decir, comunidades de red. Existen diversos algoritmos capaces de detectar comunidades de nodos en una red; unos son más adecuados para redes grandes, otros para redes con enlaces dirigidos, otros tienen en cuenta el peso de los enlaces (Plantíe y Plantíe, 2013). Así, el algoritmo conocido como Infomap (Bohlin y otros, 2014), clasifica los nodos en un número relativamente elevado de comunidades; pero de ellas sólo unas pocas son significativas por su tamaño. El resto son comunidades de muy pocos miembros, por lo general uno o dos descriptores. La Figura 5 refleja esta distribución de comunidades.

internet	273.0
social_media	233.0
knowledge_management	230.0
library	202.0
information	198.0
academic_library	156.0
bibliometrics	155.0
information_retrieval	153.0
digital_library	150.0
research	127.0
knowledge	121.0
e_government	119.0
information_technology	111.0
information_literacy	100.0
citation_analysis	87.0
open_access	87.0
collaboration	86.0
innovation	86.0
trust	83.0
knowledge_sharing	83.0

Tabla III. Palabras clave con mayor *strength*

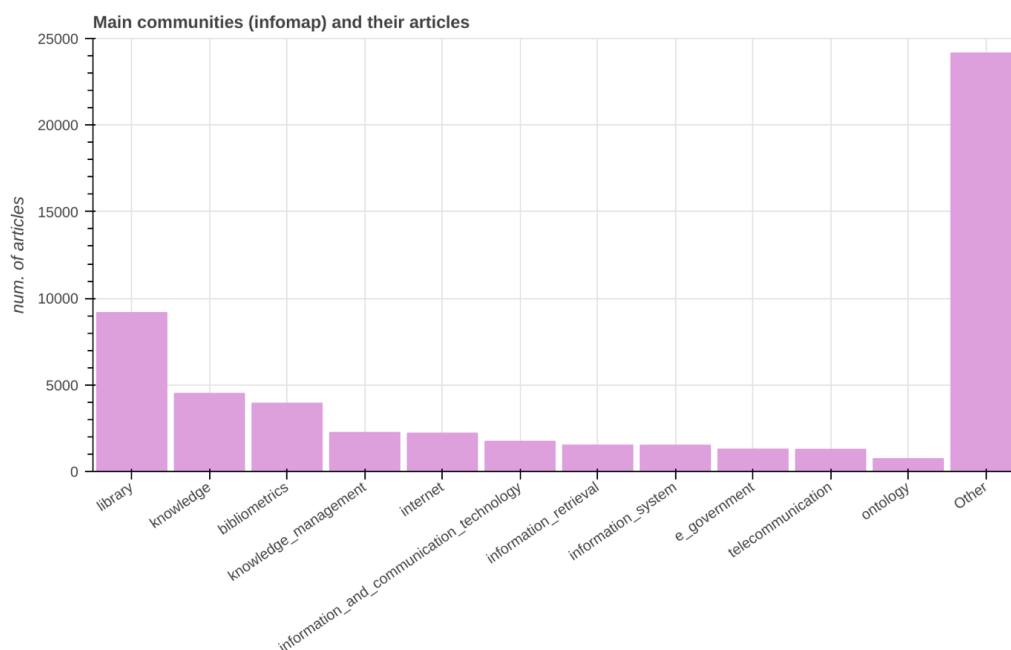


Figura 5. Comunidades Infomap

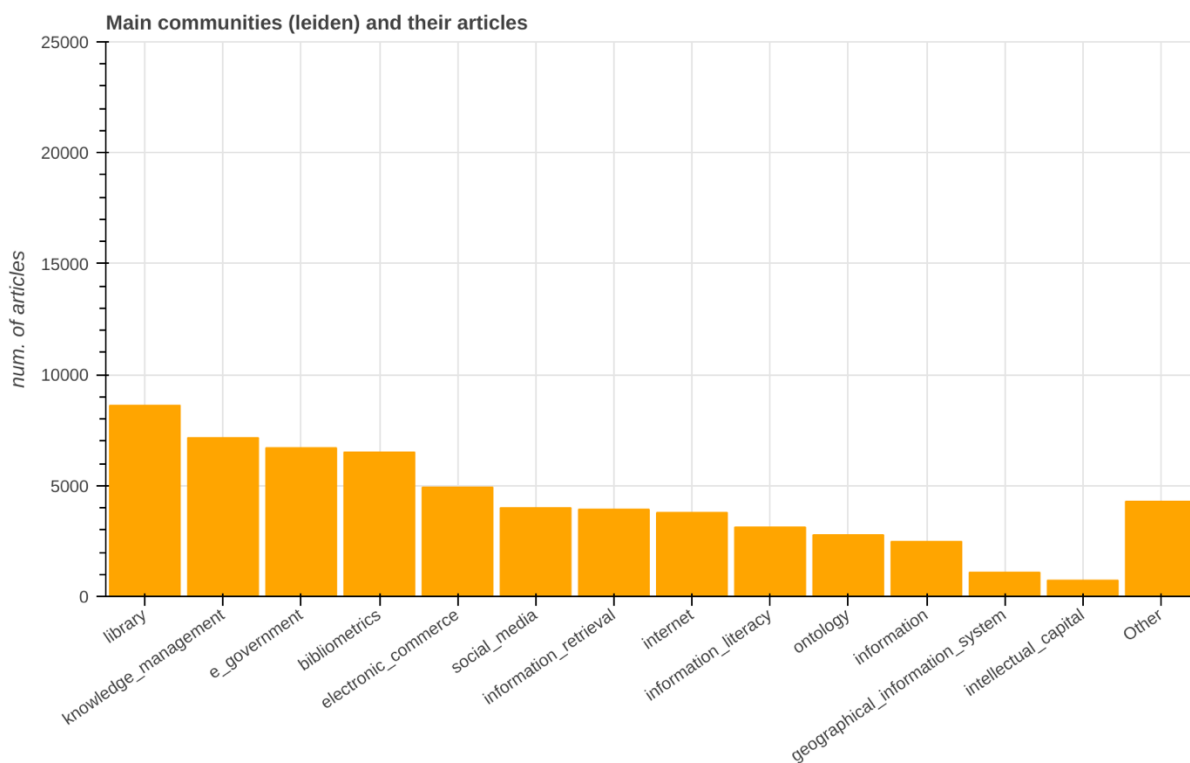


Figura 6. Comunidades Leiden

Por el contrario, el algoritmo de Leiden (Traag, Waltman y Van Eck, 2019) encuentra pocas comunidades pero con muchos miembros (Figura 6). La diferencia estriba en que Infomap clasifica con mayor precisión, pero con el peaje de dejar fuera (comunidades pequeñas) un número importante de descriptores, mientras que Leiden clasifica más cantidad de palabras clave pero con una precisión menos definida.

Las comunidades carecen de nombre (los algoritmos suelen asignar un número correlativo a cada una); pero aquí se ha asignado como etiqueta o nombre la palabra clave con mayor *strength* dentro de cada comunidad. No siempre esa palabra clave es lo suficientemente descriptiva sobre el contenido.

Las comunidades más importantes y sus descriptores más significativos pueden encontrarse en la Tabla IV.

<b>library</b>	digital_library academic_library interlending worldwide_web document_delivery electronic_book
<b>knowledge_management</b>	knowledge information_system knowledge_sharin case_study outsourcing knowledge_transfer
<b>e_government</b>	information_technology information_and_communication_technology telecommunication digital_divide broadband
<b>bibliometrics</b>	citation_analysis citation h_index open_access research_impact_factor

**electronic\_commerce** trust privacy technology\_acceptance\_model culture ethic security supply\_chain

**social\_media** web\_2.0 social\_network collaboration network blog facebook twitter

**information\_retrieval** search\_engine user\_interface world\_wide\_web user design precision text\_mining

**internet** cd\_rom web\_survey standard computer\_mediated\_communication expert\_system online\_searching

**information\_literacy** e\_learning education multimedia training learning information\_seeking distance\_learning

**ontology** metadata evaluation semantic\_web classification quality preservation

**information** implementation decision\_support\_system enterprise\_resource\_planning\_system information\_quality

**geographical\_information\_system** open\_source\_software data software\_development uncertainty digital\_elevation\_model map

**intellectual\_capital** data\_mining business\_intelligence decision\_support data\_warehouse real\_option

other

Tabla IV. Comunidades de palabras clave

Algunas comunidades aparecen muy bien perfiladas, y se corresponden con campos temáticos bien definidos, con metodología muy específica. Un caso claro es el de *bibliometrics* y otros descriptores claramente relacionados con éste. Esta es una comunidad de palabras clave de pequeño desarrollo hasta el año 2000, más o menos; de

hecho, el propio descriptor *bibliometrics* es muy poco utilizado (Figura 7). Es a partir de 2008 ó 2009 cuando este bloque de descriptores crece y se expande (Figura 8).

Es interesante, por otra parte, la relación entre unas comunidades o subcampos temáticos con

otras. Así, el grupo de descriptores en torno a *ontology* (*semantic\_web*, etc.) conecta con el de *library*; lo hace sobre todo a través de dos palabras clave: *evaluation* y *metadata*. Esto sucede a partir de 2003, de forma tímida y ya claramente a partir de 2007 (Figura 9).

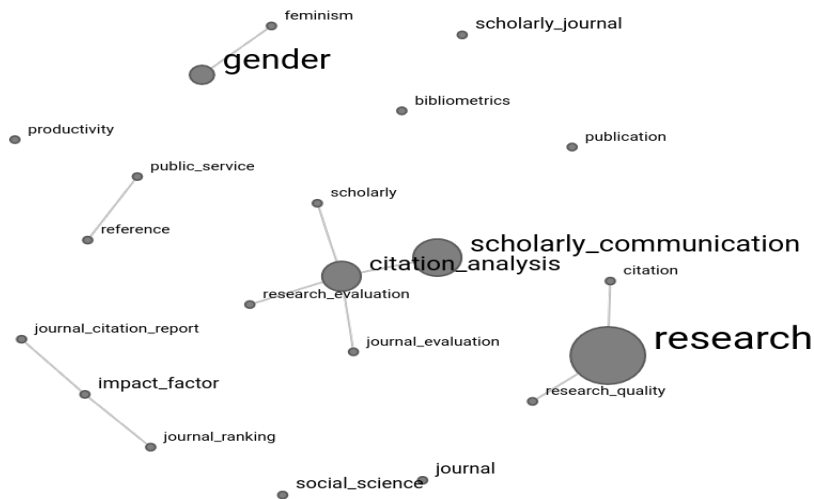


Figura 7. Comunidad de palabras clave bibliometrics en 2000

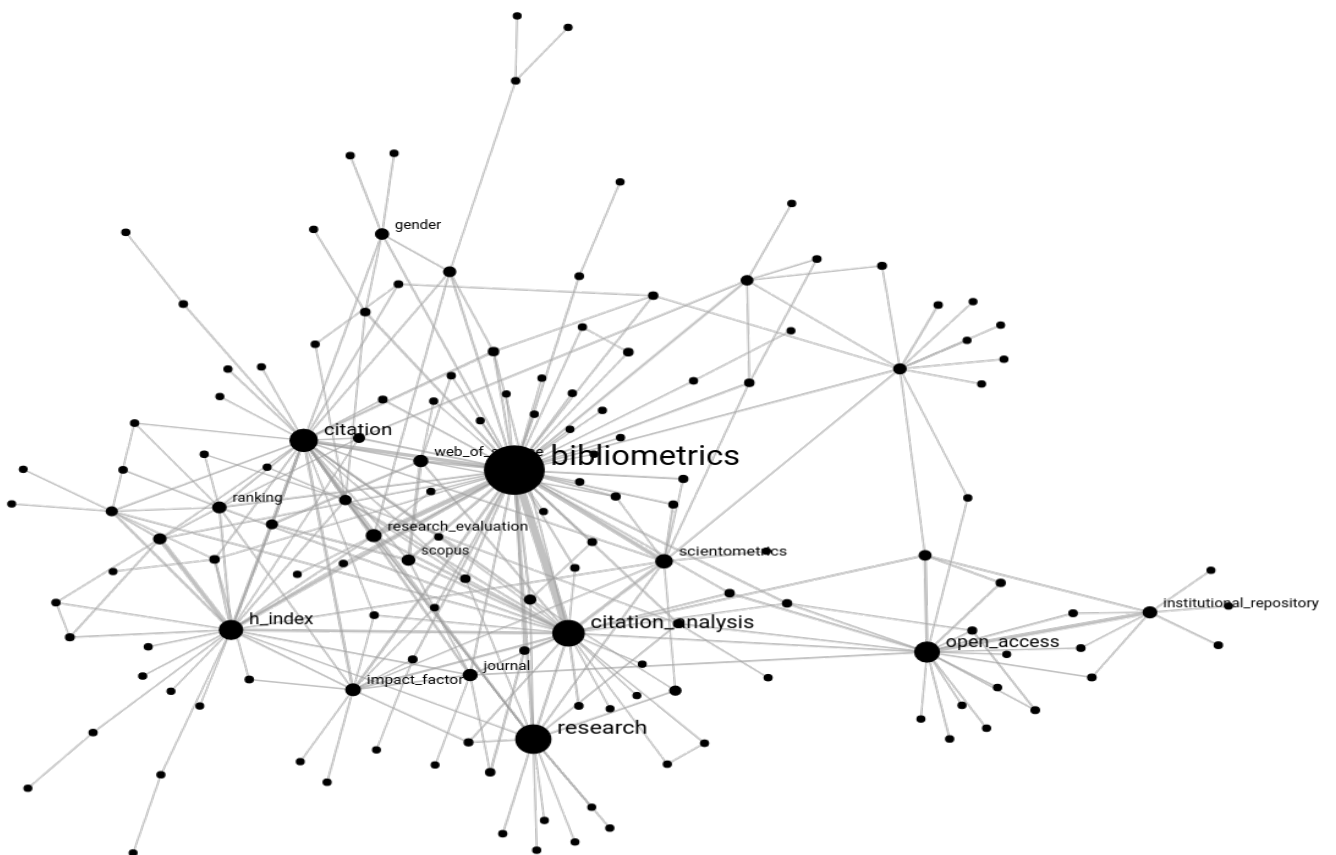


Figura 8. Comunidad bibliometrics en 2020

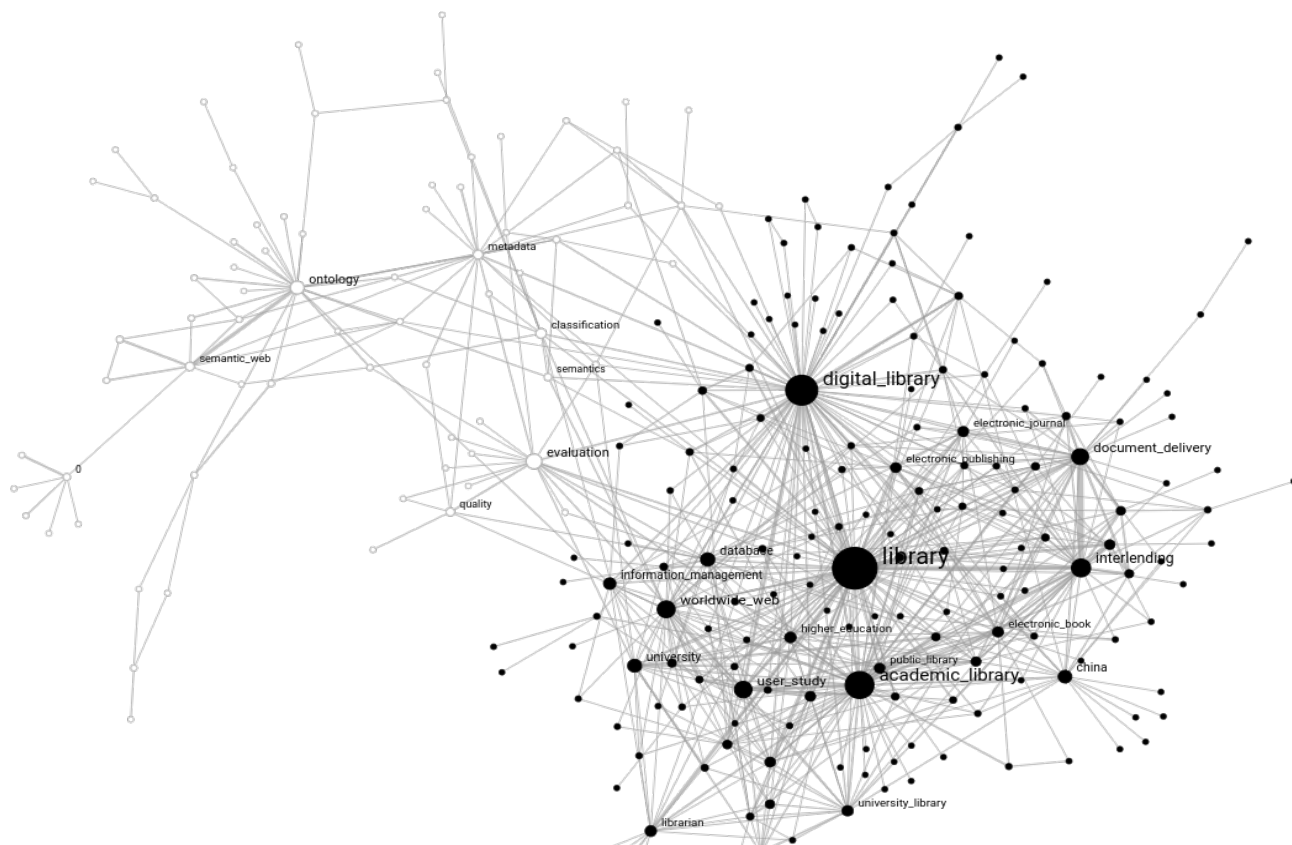


Figura 9. Comunidades Ontology y Library (2020)

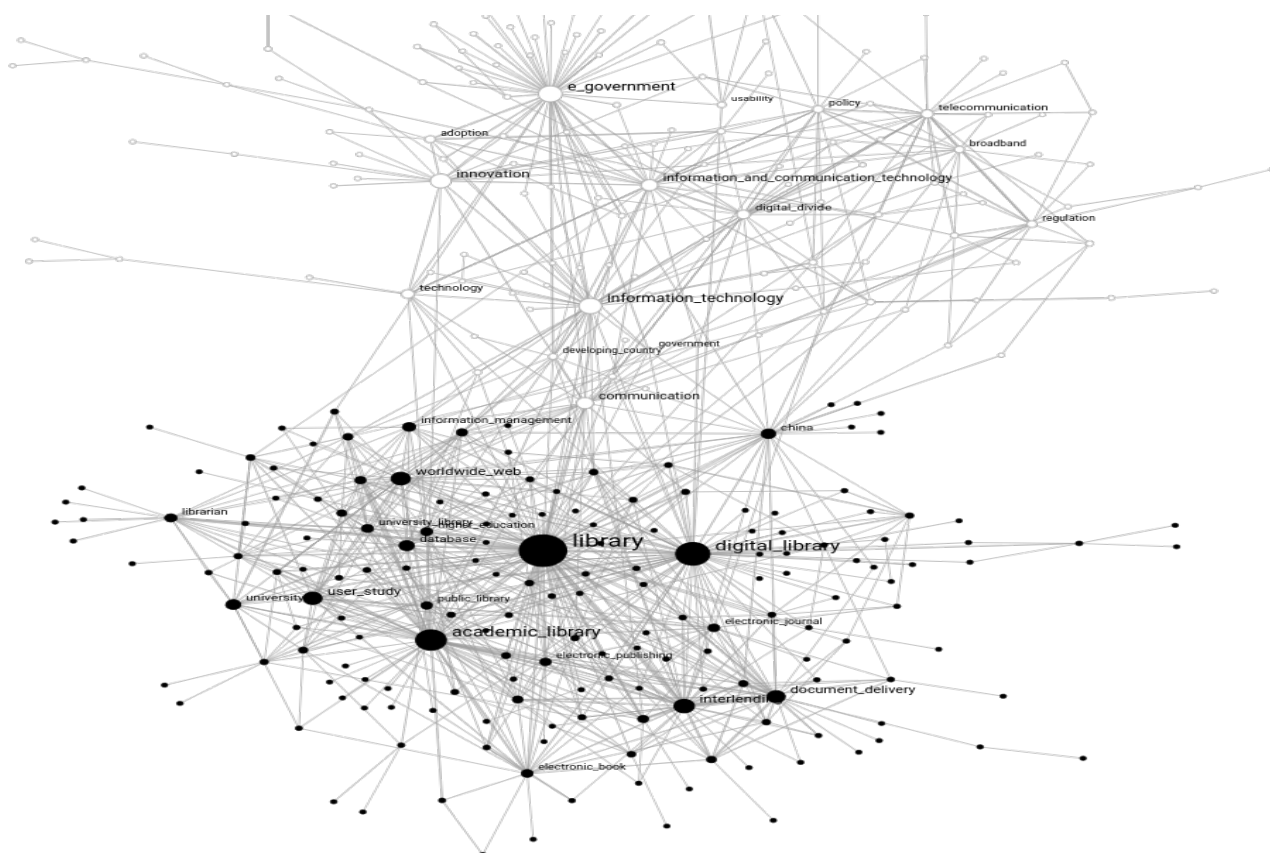


Figura 10. Comunidades e-government y library

Los bloques *library* y *e\_government* tienen poca relación entre sí; la que tienen, sin embargo, se produce a través de descriptores como *telecommunication* y *communication*, pero también a través de dos descriptores que introducen un factor regional: *china* e *india* (Figura 10, en la página anterior). La palabra clave *electronic\_commerce* y descriptores coincidentes forman una comunidad

específica, donde, además de descriptores relacionados directamente con el comercio electrónico, encontramos otros como *trust*, *privacy*, *security*, pero también *ethic* y *censorship*. La conexión de esta comunidad con la de *library* es interesante, pues se produce a través de descriptores como *world\_wide\_web* y *user\_study*, pero también a través de *china* (Figura 11).

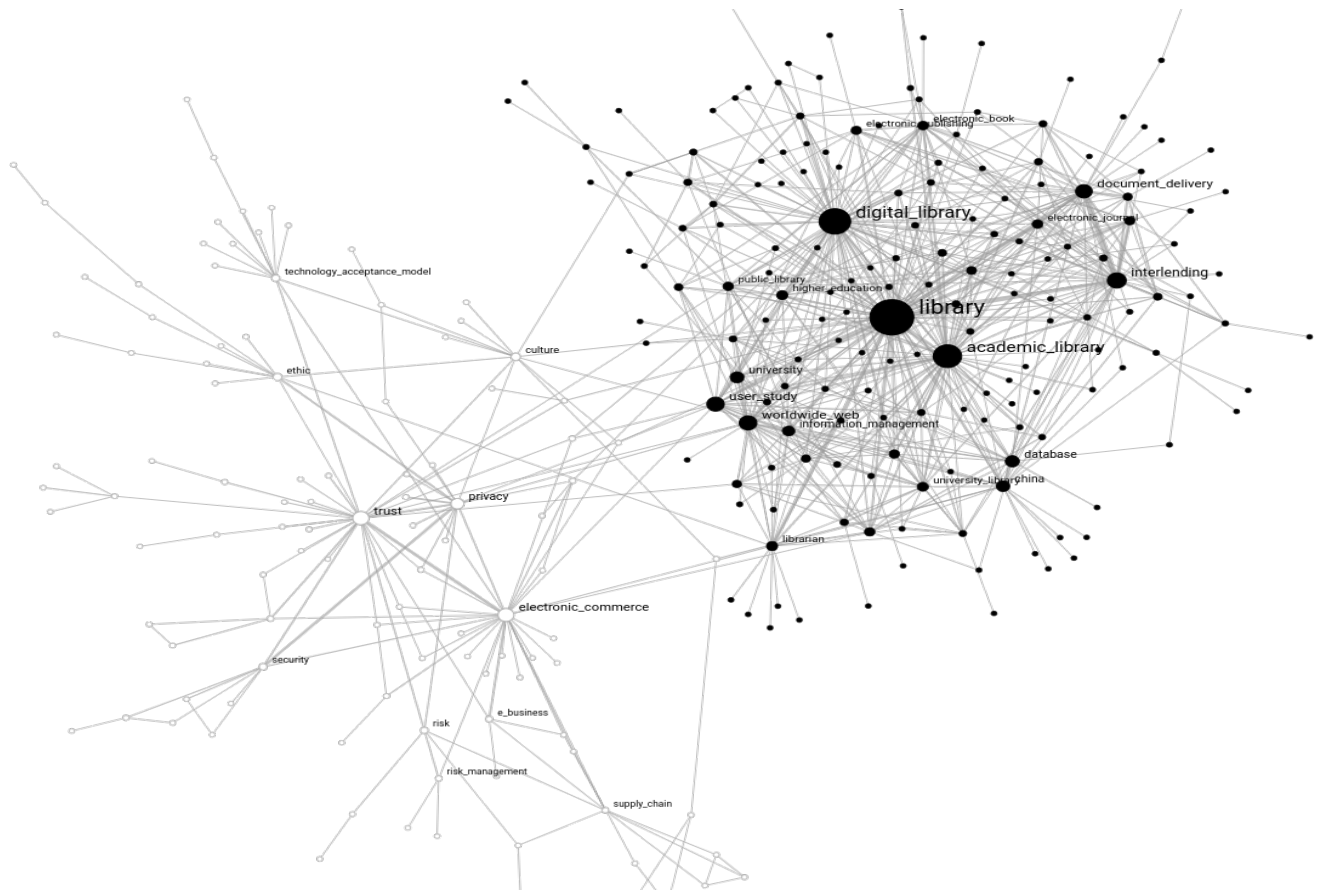


Figura 11. *library* y *e\_commerce*

La comunidad de *information\_literacy* tiene un descriptor destacado, *multimedia*, que surge relativamente pronto (sobre los años 90 del s. XX) pero que muestra un cierto declive en favor de otros como *e\_learning*. Conecta con el grupo de *library* a través de palabras clave obvias, como *university* o *academic\_library*; pero también a través de *database* o *user\_study* (Figura 12, en la página siguiente).

Por otra parte, *knowledge\_management* es interesante, porque en sus inicios (hasta 2001) (Figura 13) el descriptor dominante en su grupo es *information\_system*, muy ligado a aspectos relacionados con la ingeniería y con la información en las organizaciones. La situación cambia rápi-

damente y en 2010 las palabras clave relacionadas con el conocimiento (*management*, *sharing* y otros) pasan a ser las palabras clave más relevantes en ese grupo (Figura 14).

## 5. Conclusiones

Los descriptores o palabras clave proporcionados por los autores de artículos académicos proporcionan una visión de los contenidos temáticos que puede ayudar a conocer mejor los contenidos de campos de investigación científica y su evolución. Es posible construir redes de descriptores basadas en la coincidencia de éstos, es decir, de su uso conjunto en los mismos artículos. De otro lado, la inclusión en estas redes de la fecha de publicación, permite dotar a dichas redes de un factor dinámico.



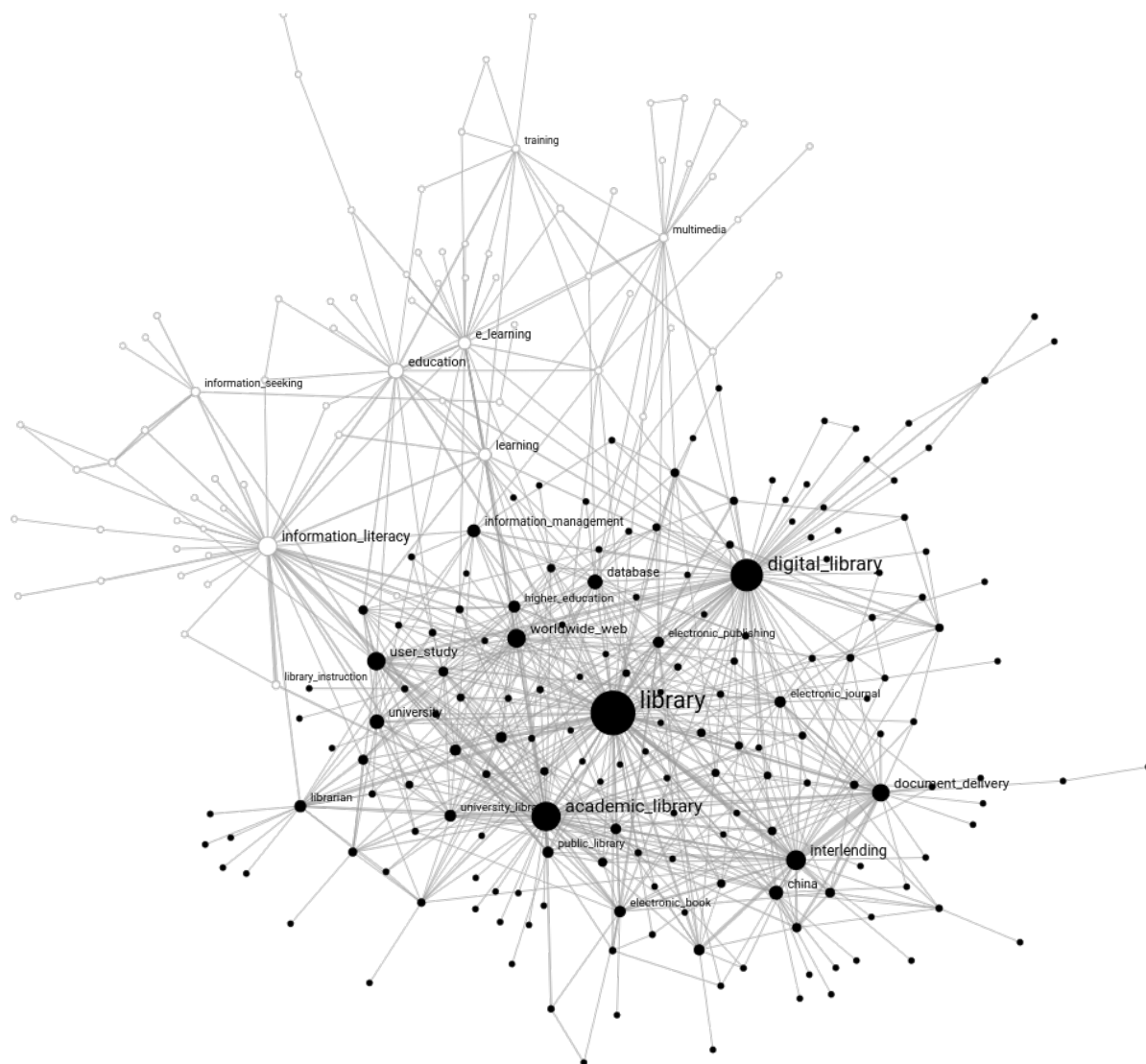


Figura 12. Library e information\_literacy

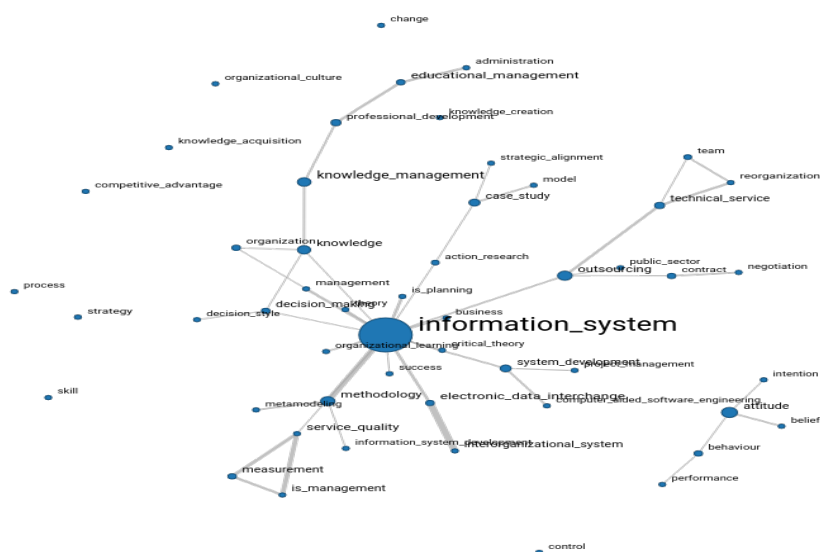


Figura 13. Comunidad knowledge\_management en 2001

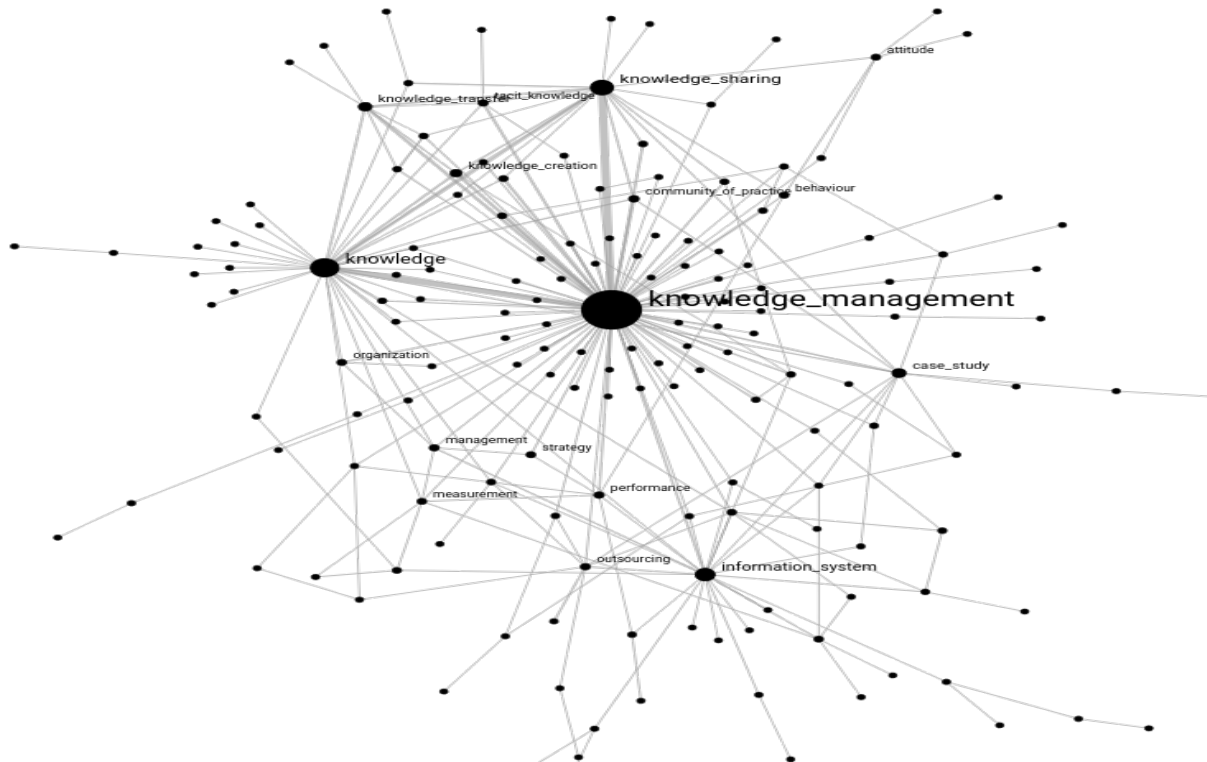


Figura 14. *knowledge\_management* en 2010

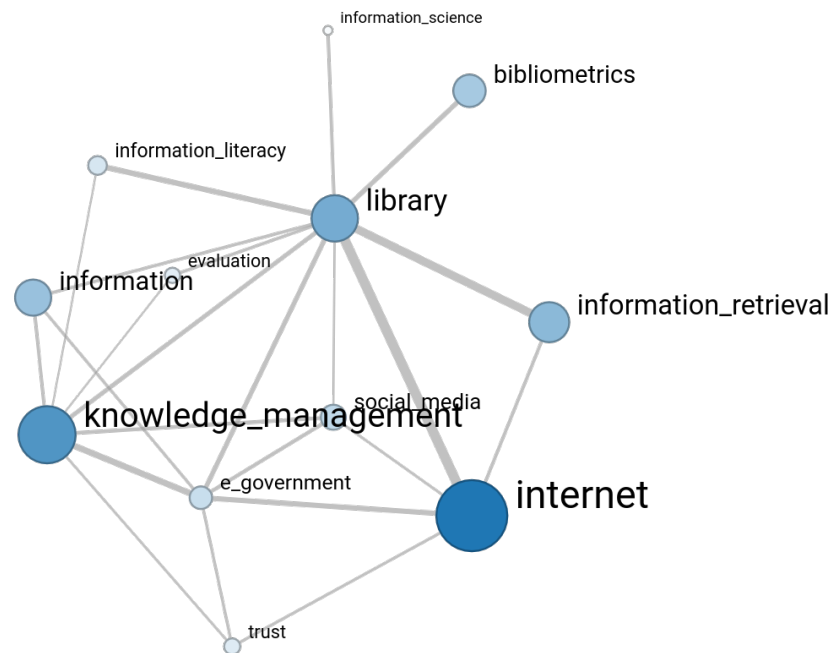


Figura 15. Red de comunidades de descriptores (1971-2020)

El análisis de estas redes mediante instrumentos adecuados como el software *netCoin*, permite observar las relaciones estadísticamente más notables entre las palabras clave y no sólo la mera frecuencia de cada una de ellas individualmente. Además, permite observar la evolución

temporal de grupos de descriptores que se utilizan con frecuencia juntos.

Así, en el campo de *Library and Information Science* se han podido encontrar 13 grupos o comunidades importantes de palabras clave (Fi-

gura 15, en la página anterior). Cada uno de estos grupos consta de varios descriptores que se utilizan de forma conjunta repetidamente. Algunos parecen autocontenidos, dibujando subcampos temáticos bien definidos, como *bibliometrics*. Otros conectan de manera abundante con palabras clave de otros subcampos; es el caso de *library* o *internet*, que funcionan en gran medida como comodines genéricos que, con el paso del tiempo, van dejando su lugar en favor de palabras clave más especializadas.

En conclusión, las técnicas de análisis de redes, junto con los instrumentos adecuados para aplicar estas técnicas, pueden resultar de gran ayuda para profundizar en la estructura temática de un dominio de conocimiento y en su evolución.

## Agradecimientos

Este trabajo ha sido financiado por Ministerio de Ciencia, Innovación y Universidades, Programa estatal de Generación del Conocimiento. Ref.: PGC2018-093755-B-I00.

## 6. Referencias

- Abrizah, A.; Zainab, A. N.; Kiran, K.; Raj, R. G. (2013). LIS journals scientific impact and subject categorization: a comparison between Web of Science and Scopus. // *Scientometrics*. 94:2, 721-740
- Arif, T. (2015). The mathematics of social network analysis: Metrics for academic social networks. *International Journal of Computer Applications Technology and Research*. 4:12, 889-893.
- Bohlin, L.; Edler, D.; Lancichinetti, A.; Rosvall, M. (2014). Community detection and visualization of networks with the map equation framework. // *Measuring scholarly impact*. Springer. 3-34
- Chen, X.; Chen, J.; Wu, D.; Xie, Y.; Li, J. (2016). Mapping the research trends by co-word analysis based on keywords from funded project. // *Procedia Computer Science*. 91, 547-555.
- Cheng, F.-F.; Huang, Y.-W.; Yu, H.-C.; Wu, C.-S. (2018). Mapping knowledge structure by keyword co-occurrence and social network analysis: Evidence from Library Hi Tech between 2006 and 2017. // *Library Hi Tech*. 36:4. <https://doi.org/10.1108/LHT-01-2018-0004>
- Csardi, G.; Nepusz, T. (2006). The igraph software package for complex network research. // *Inter. Journal, Complex Systems*, 1695.
- Doreian, P.; Woodard, K. L. (1994). Defining and locating cores and boundaries of social networks. // *Social networks*. 16:4, 267-293.
- Dotsika, F.; Watkins, A. (2017). Identifying potentially disruptive trends by means of keyword network analysis. // *Technological Forecasting and Social Change*. 119, 114-127.
- Duvvuru, A.; Kamarthi, S.; Sultornsanee, S. (2012). Undercovert research trends: Network analysis of keywords in scholarly articles. // 2012 Ninth International Conference on Computer Science and Software Engineering (JCSSE). 265-270.
- Escobar, M. (2015). Studying Coincidences with Network Analysis and other Multivariate Tools. // *The Stata Journal*. 15:4, 1118-1156.
- Escobar, M.; Martínez, L. (2020). Network Coincidence Analysis: the netCoin R Package. // *Journal of Statistical Software*. 93:11, 1-31.
- Figueroa, C. G.; García Marco, J. & Pinto, M. (2017). Mapping the evolution of library and information science (1978–2014) using topic modeling on LISA. // *Scientometrics*. 112:3 1507-1535. <https://doi.org/10.1007/s11192-017-2432-9>
- Gil Leiva, I.; Alonso Arroyo, A. (2005). La relación entre las palabras clave aportadas por autores de artículos de revista y su indización en las bases de datos ISOC, IME e ICYT.
- Gil-Leiva, I.; Rodríguez-Muñoz, J.-V. (1997). Análisis de los descriptores de diferentes áreas de conocimiento. // *Revista española de documentación científica*. 20:2, 150-160.
- Jaewoo, C.; Woonsun, K. (2014). Themes and trends in Korean educational technology research: A social network analysis of keywords. // *Procedia-Social and Behavioral Sciences*. 131, 171-176.
- Leydesdorff, L. (2002). The communication turn in the theory of social systems. // *Systems Research and Behavioral Science: The Official Journal of the International Federation for Systems Research*. 19:2, 129-136.
- Leydesdorff, L.; Vaughan, L. (2006). Co-occurrence matrices and their applications in information science: Extending ACA to the Web environment. // *Journal of the American Society for Information Science and technology*. 57(12), 1616-1628.
- Lee, P.-C.; Su, H.-N. (2010). Investigating the structure of regional innovation system research through keyword co-occurrence and social network analysis. // *Innovation*. 12:1, 26-40.
- Leiva, I. G. (2002). Consistencia en la indización de documentos entre indizadores noveles. // *Anales de documentación*. 5, 99-111.
- Li, M. (2018). Classifying and ranking topic terms based on a novel approach: Role differentiation of author keywords. // *Scientometrics*. 116:1, 77-100. <https://doi.org/10.1007/s11192-018-2741-7>
- Lozano, S.; Calzada-Infante, L.; Adenso-Díaz, B.; García, S. (2019). Complex network analysis of keywords co-occurrence in the recent efficiency analysis literature. // *Scientometrics*. 120:2, 609-629.
- Lu, W.; Liu, Z.; Huang, Y.; Bu, Y.; Li, X.; Cheng, Q. (2020). How do authors select keywords? A preliminary study of author keyword selection behavior. // *Journal of Informetrics*. 14:4, 101066.
- Miller, George A. (1995). WordNet: A Lexical Database for English. // *Communications of the ACM*. 38:11: 39-41.
- Min, K.; Yoon, M.; Furuya, K. (2019). A Comparison of a smart city's trends in urban planning before and after 2016 through keyword network analysis. // *Sustainability*. 11(11), 3155.
- Mokhtarpour, R.; Khasseh, A. A. (2020). Twenty-six years of LIS research focus and hot spots, 1990–2016: A co-word analysis. // *Journal of Information Science*. 0165551520932119.
- Onyanacha, O. B. (2018). Forty-five years of LIS research evolution, 1971–2015: An informetrics study of the author-supplied keywords. // *Publishing research quarterly*. 34:3, 456-470.
- Otte, E.; Rousseau, R. (2002). Social network analysis: A powerful strategy, also for the information sciences. // *Journal of information Science*. 28:6, 441-453.
- Park, J.; Jeong, E. (2019). Service quality in tourism: A systematic literature review and keyword network analysis. // *Sustainability*. 11(13), 3665.

- Park, J. S.; Kim, N. R.; Han, E. J. (2018). Analysis of trends in science and technology using keyword network analysis. // *Journal of the Korea Industrial Information Systems Research*, 23:2, 63-73.
- Peña, L. J. M. (2012). Análisis bibliométrico sobre la producción científica archivística en la Red de Revistas Científicas de América Latina y el Caribe (Redalyc) durante el período 2001-2011. // *Biblios*. 48, 1-11.
- Pérez, V. A.; Urbáez, M. F. (2016). Modelos teóricos de gestión del conocimiento: Descriptores, conceptualizaciones y enfoques. // *Entreciencias: diálogos en la Sociedad del Conocimiento*. 4:10 201-227.
- Peset, F.; Garzón-Farínos, F.; González, L. M.; García-Massó, X.; Ferrer-Sapena, A.; Toca-Herrera, J. L.; Sánchez-Pérez, E. A. (2020). Survival analysis of author keywords: An application to the library and information sciences area. // *Journal of the Association for Information Science and Technology*. 71:4, 462-473.
- Plantí, M.; Crampes, M. (2013). Survey on social community detection. // *Social media retrieval* (pp. 65-85). Springer.
- Rodríguez Cruz, Y.; Martínez Rodríguez, A. (2009). Comportamiento de la producción científica sobre gestión de información en revistas del Web of Science (1995-2008). // *Acimed*. 20:6, 101-124.
- Rosvall, M.; Axelsson, D.; Bergstrom, C. T. (2009). The map equation. // *The European Physical Journal Special Topics*. 178:1, 13-23.
- Strader, C. R. (2011). Author-assigned keywords versus Library of Congress subject headings. // *Library resources & technical services*. 53:4, 243-250.
- Tous, M. G.; Mattar, S. (2012). Las claves de las palabras clave en los artículos científicos. // *Revista MVZ Córdoba*. 17:2, 2955-2956.
- Traag, V. A.; Waltman, L.; Van Eck, N. J. (2019). From Louvain to Leiden: Guaranteeing well-connected communities. // *Scientific reports*. 9:1, 1-12.
- Tripathi, M.; Kumar, S.; Sonker, S. K.; Babbar, P. (2018). Occurrence of author keywords and keywords plus in social sciences and humanities research: A preliminary study. // *COLLNET Journal of Scientometrics and Information Management*. 12:2, 215-232.
- Uddin, S.; Khan, A. (2016). The impact of author-selected keywords on citation counts. *Journal of Informetrics*. 10:4, 1166-1177.
- Van der Hulst, R. C. (2009). Introduction to Social Network Analysis (SNA) as an investigative tool. // *Trends in Organized Crime*. 12:2, 101-121.
- Woolgar, S. and Latour, B. (1988) *La vie de laboratoire: la production des faits scientifiques*. Editions La Découverte, 1988.
- Yoo, S.; Jang, S.; Byun, S. W.; Park, S. (2019). Exploring human resource development research themes: A keyword network analysis. // *Human Resource Development Quarterly*. 30:2, 155-174.

---

Enviado: 2021-05-12. Segunda versión: 2021-06-21.

Acceptado: 2021-06-21.

---

---

# Índice de autores

*Author index*

---

Agenjo Bullón, Xavier, 13  
Akerreta Escribano, Lorea, 31  
Alonso Berrocal, José Luis, 71  
Díaz-Corona, Dayany, 23  
Escobar Mercado, Modesto, 71  
Figueroa, Carlos G., 71  
Fujita, Mariângela Spotti Lopes, 45  
García Marco, Francisco Javier, 55

Hernández Carrascal, Francisca, 13  
Lacasta, Javier, 23  
Moyano Collado, Julián, 31  
Nogueras-Iso, Javier, 23  
Silva, Bruna Daniele de Oliveira, 45  
Tolare, Jessica Beatriz, 45  
Zazo Rodríguez, Ángel, 71

---

# Índice de materias en español

*Subject index in Spanish*

---

Aragón, 31  
Bibliometría, 71  
Biblioteconomía y documentación, 71  
Data storytelling, 31  
Datos abiertos, 13, 31  
Datos abiertos de patrimonio cultural, 13  
Descriptor, 71  
Detección de comunidades, 71  
Europeana, 13  
Folksonomías, 45  
Goodreads, 45  
Indexación social, 45  
Innovación pública, 31  
Metadatos, 23

Modelos de organización del conocimiento, 23  
NetCoin, 71  
Palabras clave, 71  
Patrimonio cultural digitalizado, 13  
Redes dinámicas, 71  
Representación y recuperación de información, 45  
Reutilización de la información, 31  
Sistemas de organización del conocimiento, 23  
Skoob, 45  
SKOS, 23  
Técnicas de análisis de redes, 71  
Transparencia, 31  
Turismo cultural, 13  
Web semántica, 23

---

# Índice de materias en inglés

*Subject index in English*

---

Bibliometrics, 71  
Community detection, 71  
Cultural heritage, 13  
Cultural tourism, 13  
Data storytelling, 31  
Descriptors, 71  
Digital cultural heritage, 13  
Europeana, 13  
Folksonomy, 45  
Goodreads, 45  
Information representation, 45  
Information retrieval, 45  
Information reuse, 31  
Keywords, 71

Knowledge organization systems, 23  
Library and information science, 71  
Linked Open Data, 13  
Metadata, 23  
NetCoin, 71  
Network analysis techniques, 71  
Open data, 31  
Public innovation, 31  
Recommendation systems, 23  
Semantic web, 23  
Skoob, 45  
SKOS, 23  
Social indexing, 45  
Transparency, 31