

Item

4 Editorial

6 Dossier

Entorns de cerca i ús de la informació

- 7 Tecnologies de dades enllaçades (*linked data*). El que les biblioteques han aconseguit fins ara
Yongming WANG; Sharon Q. YANG

- 28 Evolució de les dades obertes enllaçades (LOD) en el context dels vocabularis controlats
Rosa FABIRO; Carme MASAGUÉ; Anna ROVIRA

- 50 Les Google Accelerated Mobile Pages: la velocitat de càrrega com a factor de posicionament SEO
Rubén ALCARAZ

- 75 Classificació semàntica i visual de documents digitals
Marçal RUSIÑOL

89 Caixa d'eines

L'ús de les dades d'acord amb el Reglament General de Protecció de Dades
Eloi FONT

100 Retrovisor

Les marques de propietat dels llibres de la Biblioteca d'Escornalbou: un patrimoni recuperat
Pedro RUEDA; Dolors SAUMELL



Item: revista de biblioteconomia i documentació
Núm. 65-66 - Barcelona - juliol-desembre 2018 / gener-juny 2019
Editada pel Col·legi Oficial de Bibliotecaris-Documentalistes de Catalunya
ISSN 0214-0349 (paper) · ISSN 1699-521X (Internet)

Editorial

Aquest número inclou algunes novetats. La primera, i més evident, és la del canvi de format. Al darrere d'aquest nou disseny hi ha la voluntat de la Junta del Col·legi de Bibliotecaris-Documentalistes de Catalunya i del Consell Editorial de la revista d'oferir-vos una publicació més manejable, més propera als vostres interessos, amb imatge i continguts nous, que reflecteixi el dinamisme i les capacitats d'adaptació i d'innovació propis del nostre àmbit professional.

El nou *Ítem* parteix també de la voluntat d'integrar un ampli ventall de punts de vista, tal com s'ha estat fent, i recollir les noves tendències. Iniciem una nova etapa de la revista amb un Consell Editorial renovat, que beu de la tradició i agraeix des d'aquí la feina i la implicació de tots els membres que han fet possible la revista en els darrers anys.

Una altra característica d'aquest número, en aquest cas excepcional, és que es tracta d'un número doble —65/66 (juliol-desembre 2018 i gener-juny 2019). Es publica d'aquesta manera per poder avançar la sortida de cada número a mitjan semestre, com a màxim, mantenint-ne la periodicitat.

Hem volgut iniciar una nova etapa oferint els avenços en els entorns de cerca i ús d'informació, les tecnologies de dades enllaçades (*linked data*) i el web semàntic. Parlar avui dia de cerca d'informació és, per a la majoria de la població amb connexió a internet, parlar de Google; amb més de 6 bilions de cerques diàries arreu del món,¹ representa un 65 % de les cerques al web. En el nostre entorn professional, parlar de cerca d'informació és parlar també de catàlegs, eines de descoberta, llenguatges contro-

lats, organització del coneixement, índexs de recuperació de la informació, etc.

El web semàntic i les dades obertes enllaçades ens facilitaràn l'accés a informació de qualitat i als resultats més rellevants per als usuaris, fent visible als algorismes dels cercadors web, cada cop més, el treball dut a terme per les biblioteques d'arreu del món. Sabem que el camí per aconseguir la interoperabilitat dels sistemes no és fàcil; malgrat que hi hagi voluntat dels professionals, la major part d'empreses que comercialitzen les plataformes de gestió d'informació i biblioteques no han prioritzat aquest factor perquè existeixen múltiples models conceptuals,² entre altres raons, i no existeix un estàndard únic que s'hagi imposat. Així i tot, són nombroses les biblioteques i els portals que han engegat projectes vinculats a l'àmbit.

Pel que fa a l'ús de la informació, la realitat web evoluciona a gran velocitat i l'aplicació de la intel·ligència artificial ha permès tant el tractament automatitzat de grans volums de dades com també l'aprenentatge automàtic. Aquests avenços són molt rellevants per a la investigació en sectors com el de la salut i el disseny de serveis amb valor afegit, i on han trobat un gran aliat ha estat en el comerç digital. En aquest punt, han sorgit dubtes sobre les pràctiques de les grans empreses i l'obtenció i l'ús que fan de les dades que disposen de la ciutadania.³ Motius suficients a la Comissió Europea, que a mitjan desembre de 2018 va donar a conèixer l'esborrany del Projecte de directrius ètiques per a una IA confiable,⁴ la versió final del qual s'espera que es faci públic l'abril de 2019. Aquest document té com a objectiu establir les bases del desenvolupament de la intel·ligència artificial des d'un punt de vista ètic i al servei del benestar de les persones.

1. < <http://www.internetlivestats.com/>>.

2. Sofia Zapounidou, Michalis Sfakakis, Christos Papatheodorou, «Representing and integrating bibliographic information into the semantic web: a comparison of four conceptual models», *Journal of Information Science*, v. 43, n. 4 (2017), p. 525-553.

3. David Casacuberta, «Màquines socials i internet: hi va haver algun problema pel camí» [en línia]. *BID: textos universitaris de biblioteconomia i documentació*, n. 41 (2018), <<http://bid.ub.edu/41/casacuberta.htm>> [Consulta: 14/01/2019].

4. Comissió Europea, «Draft Ethics guidelines for trustworthy AI, Digital Single Market, High-Level Expert Group on Artificial Intelligence (AI HLEG), Report / Study, 18 Dec. 2018» [en línia], [Brussels: European Commission, 2018], <<https://ec.europa.eu/digital-single-market/en/news/draft-ethics-guidelines-trustworthy-ai>> [Consulta: 14/01/2019].



El número que teniu a les mans recull al **Dossier** monogràfic quatre contribucions sobre les tecnologies de dades enllaçades, el web semàntic i els projectes per millorar l'experiència de l'usuari en la recuperació de la informació.

En primer lloc, trobareu la traducció d'un article de dos autors estatunidencs, Yongming Wang i Sharon Q. Yang, publicat originalment a l'*International Journal of Librarianship* i que analitza/tracta l'evolució internacional de les iniciatives de dades enllaçades que s'han dut a terme a les biblioteques en els darrers vint anys.

En segon lloc, Rosa Fabeiro, Carme Masagué i Anna Rovira fan un recull dels projectes que apliquen les dades obertes enllaçades (LOD) en el nostre àmbit, especialment dels projectes relacionats amb els llenguatges controlats d'indexació i classificació, i exposen l'experiència aplicada al Thesaurus de la Universitat de Barcelona.

A continuació, Rubén Alcaraz analitza la iniciativa Accelerated Mobile Pages de Google, que facilita la creació de pàgines web que es carreguen molt ràpidament en qualsevol plataforma de consulta, i d'aquesta manera aconseguir-

en una clara millora en el posicionament natural de la pàgina en motors de cerca.

L'article que tanca el dossier és de Marçal Rusiñol que presenta com s'aplica la intel·ligència artificial i els algorismes d'aprenentatge profund en la classificació i la cerca de documents digitals.

La secció **Caixa d'eines** inclou una contribució d'Eloi Font sobre l'ús de les dades d'acord amb el nou Reglament General de Protecció de Dades (RGPD).

Tanca el número la secció **Retrovisor**, amb un article de Pedro Rueda i Dolors Saumell dedicat al patrimoni bibliogràfic d'Escornalbou, i, més concretament, a les marques de propietat dels llibres de la biblioteca del convent.

Mirades complementàries, innovadores, procedents d'àmbits i entorns molt diversos i, fins i tot, contraposades en el temps, però alineades en la identificació i l'explotació de les dades que han de permetre la millora dels processos de cerca i ús de la informació i dels sistemes d'organització del coneixement.

El Consell Editorial ■



Yongming WANG

Library, The College of New Jersey, Nova Jersey, EUA
wangyo@tcnj.edu

Sharon Q. YANG

Moore Library, Rider University, Nova Jersey, EUA
yangs@rider.edu

Article traduït amb el permís dels autors:

Yongming Wang; Sharon Q. Yang «Linked data technologies and what libraries have accomplished so far» [en línia]. *International Journal of Librarianship*, v. 3, n. 1 (2018), p. 3-20.
<<https://doi.org/10.23974/ijol.2018.vol3.1.62>>
[Consulta: 28/11/2018].

Tecnologies de dades enllaçades (*linked data*): el que les biblioteques han aconseguit fins ara

Resum: En els darrers deu anys, les biblioteques han dedicat esforços en l'àmbit de les dades enllaçades i del web semàntic. Sovint costa entendre els detalls tècnics i el potencial que tenen per a la comunitat bibliotecària, a causa de la seva complexitat i ampli abast. Aquest article pretén aclarir alguns conceptes importants i explica quines són les tecnologies bàsiques de les dades enllaçades: l'RDF (marc de descripció de recursos), l'ontologia i el llenguatge de cerca. Es descriuen alguns dels assoliments obtinguts en biblioteques d'arreu del món pel que fa a la transformació de les seves dades en dades enllaçades, incloent-hi la tasca de la Library of Congress, de l'Online Computer Library Center (OCLC) i altres biblioteques nacionals. També s'analitzen alguns dels reptes i contratemps als quals s'han enfrontat les biblioteques a l'hora de desenvolupar aquestes tecnologies. Tanmateix, no es pot tornar enrere: les biblioteques ho han d'aconseguir.

Paraules clau: dades enllaçades (*linked data*), web semàntic, RDF (marc de descripció de recursos), BIBFRAME, Library of Congress, OCLC.

Tecnologías de datos enlazados (*linked data*): lo que las bibliotecas han conseguido hasta ahora

Resumen: En los últimos diez años, las bibliotecas han dedicado esfuerzos en el ámbito de los datos enlazados y la web semántica. A menudo cuesta entender los detalles técnicos y el potencial que tienen para la comunidad bibliotecaria, debido a su complejidad y amplio alcance. Este artículo pretende aclarar algunos conceptos importantes y explica cuáles son las tecnologías básicas de los datos enlazados: RDF (marco de descripción de recursos), la ontología y el lenguaje de búsqueda. Se describen algunos de los logros obtenidos en bibliotecas de todo el mundo en cuanto a la transformación de sus datos en datos enlazados, incluidas la tarea de la Library of Congress, del Online Computer Library Center (OCLC) y de otras bibliotecas nacionales. También se analizan algunos

de los retos y contratiempos a los que se han enfrentado las bibliotecas a la hora de desarrollar estas tecnologías. Sin embargo, no se puede volver atrás: las bibliotecas lo deben conseguir.

Palabras clave: datos enlazados (*linked data*), web semántica, RDF (marco de descripción de recursos), BIBFRAME, Library of Congress, OCLC.

Linked data technologies and what libraries have accomplished so far

Abstract: For the past ten years libraries have been working diligently towards Linked Data and the Semantic Web. Due to the complexity and vast scope of Linked Data, many people have a hard time to understand its technical details and its potential for the library community. This paper aims to help librarians better understand some important concepts by explaining the basic Linked Data technologies that consist of Resource Description Framework (RDF), the ontology, and the query language. It also includes an overview of the achievements by libraries around the world in their efforts to turn library data into Linked Data including those by Library of Congress, Online Computer Library Center (OCLC), and some other national libraries. Some of the challenges and setbacks that libraries have encountered are analyzed and discussed. In spite of the difficulties, there is no way to turn back. Libraries will have to succeed.

Keywords: Linked Data, Semantic Web, Resource Description Framework, BIBFRAME, Library of Congress, OCLC.



Introducció

Què són les dades enllaçades? Segons David Wood, el copresident del grup de treball de l'RDf de W3C (World Wide Web Consortium), que estableix les bases per a les dades enllaçades i el web semàntic, «són un conjunt de tècniques per representar i connectar dades estructurades al web... Les dades enllaçades converteixen el web en una base de dades mundial que anomenem *la xarxa de dades*».¹ Les tecnologies de les dades enllaçades, i en sentit més ampli, el web semàntic, han guanyat un fort impuls i popularitat al web. Les tecnologies de dades enllaçades tenen el potencial de transformar l'actual web de documents en web de dades. Imaginem-nos que en el futur a internet no només estiguessin connectats els documents del web sinó també totes les dades. I encara més, que a aquestes dades connectades hi poguessin accedir no només els humans sinó també les màquines. En altres paraules, tots els dispositius que estiguin connectats a internet podran accedir i processar aquestes dades enllaçades i, per tant, podran prendre decisions intel·ligents. Això potenciarà amb escreix la manera que tenim d'accedir a la informació i de prendre decisions informades. Aquestes idees no són noves. A finals dels anys noranta, Tim Berners-Lee,² l'inventor del web, ja tenia la visió d'un web semàntic:

Tinc un somni per al web, en què [els ordinadors] seran capaços d'analitzar totes les dades que hi ha al web: el contingut, els enllaços i les transaccions entre la gent i els ordinadors. Un web semàntic que encara s'ha de crear ho farà possible i, quan existeixi, les operacions comercials diàries, la burocràcia i les nostres vides seran gestionades per màquines que es comunicaran

amb altres màquines. Els «gestors intel·ligents» que la gent fa segles que pregona finalment seran una realitat.

L'any 2004 W3C va publicar la primera recomanació de model de dades enllaçades, l'RDf 1.0. L'any 2005 W3C va crear el Semantic Web Interest Group. I l'any 2006 Tim Berners-Lee va publicar les normes i els principis de creació de les dades enllaçades, que preparen el camí per desenvolupar i adoptar a gran escala les tecnologies de dades enllaçades.³

Els últims deu anys i escaig, s'ha apreciat que les empreses, tant petites com grans, han adoptat i utilitzat amb rapidesa les dades enllaçades. Les empreses com ara Google i Facebook fan servir les dades enllaçades per potenciar les seves connexions i la seva capacitat de cerca.⁴ L'empresa detallista BestBuy utilitza les dades enllaçades per millorar els seus resultats empresarials.⁵

1. Revisió bibliogràfica

L'any 2011 el Library Linked Data Incubator Group⁶ ja va publicar l'informe final «W3C Incubator Group Report». Aquest grup estava format per experts internacionals en els camps de la informació i de les biblioteques especialitzats en el web semàntic i les metadades. En aquest informe s'estudia la situació actual de les dades enllaçades, se'n resumeixen els usos i es fan algunes recomanacions importants per implementar-les a la comunitat bibliotecària.

Un altre esforç internacional estretament relacionat amb el web semàntic i les dades enllaçades és l'Annual International Conference on

1. David Wood, *et al.*, *Linked data: structured data on the web*, Shelter Island: Manning, 2014.
2. Tim Berners-Lee, Mark Fischetti, *Weaving the web: the original design and ultimate destiny of the World Wide Web*. San Francisco: Harper, 2000.
3. Tim Berners-Lee, «Linked Data» [en línia], [Cambridge: The World Wide Web Consortium (W3C), date: 2006/07/27, last change: 2009/06/18], <<https://www.w3.org/DesignIssues/LinkedData.html>> [Consulta: 08/02/2018].
4. David Wood, *et al.*, *op. cit.*, p. 8.
5. David Wood, *et al.*, *op. cit.*, p. 71.
6. «Library Linked Data Incubator Group Final Report» [en línia], [Cambridge: The World Wide Web Consortium (W3C), date: 2011/10/25], <<https://www.w3.org/2005/Incubator/ld/XGR-ld-20111025/>> [Consulta: 27/03/2018].

Imaginem-nos que en el futur a internet no només estiguessin connectats els documents del web sinó també totes les dades. I encara més, que a aquestes dades connectades hi poguessin accedir no només els humans sinó també les màquines.

Dublin Core and Metadata Applications, organitzada per la Dublin Core Metadata Initiative (DCMI). Aquesta conferència anual va començar l'any 1995 i només s'hi feien tallers, però l'any 2001 es va ampliar amb conferències, tutorials, presentacions i ponències revisades per experts. Des del principi, la DCMI tracta temes relacionats amb el web semàntic, especialment l'ontologia i els vocabularis. La temàtica de la conferència de 2005 és «Vocabularies in Practice», any en què una ponència a les memòries va introduir el concepte de SKOS (sistema senzill d'organització del coneixement) i va recomanar la combinació de SKOS Core i DCMI Metadata Terms.⁷ Un projecte de les memòries de la conferència de 2009 té el títol «Research on Linked Data and Co-reference Resolution» i explica la transformació d'un conjunt de dades d'autors acadèmics i les seves publicacions en dades enllaçades.⁸ Aquesta és una de les primeres publicacions que tracta de

l'aplicació de les dades enllaçades a la comunitat bibliotecària. I d'ençà de 2012, en aquest cicle de conferències anuals s'ha apreciat un interès creixent per les dades enllaçades.

A escala més pràctica, Karen Coyle⁹ va publicar «Linked Data Tools: Connecting on the Web» al *Library Technology Report*. En aquest article va introduir les tecnologies bàsiques de les dades enllaçades en format de tutorial. Un any després, Erik T. Mitchell¹⁰ va publicar «Library Linked Data: Research and Adoption» al *Library Technology Report*, que tracta de la recerca i el desenvolupament de les dades enllaçades a la comunitat bibliotecària. L'any 2016 Mitchell¹¹ va publicar un altre article que té a veure amb la pràctica i l'adopció de les dades enllaçades de mans de les biblioteques i que es titula «Library Linked Data: Early Activity and Development». Tots tres articles han tingut un paper important a l'hora de familiaritzar els bibliotecaris amb les dades enllaçades.

D'ençà de 2015, s'han publicat més articles de casos i exemples d'aplicació de dades enllaçades. L'article de Karim Tharani¹² estudia la possibilitat d'utilitzar el BIBFRAME per recollir i compartir dades bibliogràfiques com a dades enllaçades amb un estudi de cas. L'article de Jin, Hahn i Croll¹³ descriu com es volen transformar i enriquir gairebé 300.000 registres MARC de llibres electrònics en registres BIBFRAME perquè es pugui accedir més fàcilment a aquests

7. Alistair Miles, *et al.*, «SKOS core: simple knowledge organisation for the web» [en línia]. En: International Conference on Dublin Core and Metadata Applications (2005: Madrid), *DC-2005: proceedings of the International Conference on Dublin Core and Metadata Applications*, [Madrid: Universidad Carlos III, 2005], p. 3-10, <<http://dcpapers.dublincore.org/pubs/article/view/798>> [Consulta: 08/02/2018].
8. Hugh Glaser, *et al.*, «Research on linked data and co-reference resolution» [en línia]. En: International Conference on Dublin Core and Metadata Applications (2009: Seül), *DC-2009: proceedings of the International Conference on Dublin Core and Metadata Applications*, [Seoul: National Library of Korea, 2009], p. 113-117, <<http://dcpapers.dublincore.org/pubs/article/view/958/957>> [Consulta: 08/02/2018].
9. Karen Coyle, «Linked data tools: connecting on the web» [en línia], *Library Technology Reports*, v. 48, n. 4 (2012), <<http://dx.doi.org/10.5860/ltr.48n4>> [Consulta: 08/02/2018].
10. Erik T. Mitchell, «Library linked data: research and adoption» [en línia], *Library Technology Reports*, v. 49, n. 5 (2013), <<http://dx.doi.org/10.5860/ltr.49n5>> [Consulta: 08/02/2018].
11. Erik T. Mitchell, «Library linked data: early activity and development» [en línia], *Library Technology Reports*, v. 52, n. 1 (2016), <<http://dx.doi.org/10.5860/ltr.52n1>> [Consulta: 08/02/2018].
12. Karim Tharani, «Linked data in libraries: a case study of harvesting and sharing bibliographic metadata with BIBFRAME» [en línia], *Information and Library Technologies*, v. 34, n. 1 (2015), p. 5-19, <<https://doi.org/10.6017/ital.v34i1.5664>> [Consulta: 08/02/2018].
13. Qiang Jin, Jim Hahn, Gretchen Croll, «BIBFRAME transformation for enhanced discovery», *Library Resources & Technical Services*, v. 60, n. 4 (2016), p. 223-235, <<http://dx.doi.org/10.5860/lrts.60n4.223>> [Consulta: 08/02/2018].

recursos. L'article de Kimmy Szeto investiga com les dades enllaçades obertes poden transformar i potenciar la descoberta i la cerca de recursos musicals.¹⁴ Recentment, el Program for Cooperative Cataloging (PCC) de l'OCLC va portar a terme un altre projecte per transformar les metadades del llegat de les biblioteques, és a dir, els registres MARC, en dades enllaçades. La principal tasca d'aquest projecte és crear una base de dades de control d'autoritats amb els registres MARC tradicionals de gent, organitzacions i llocs de moltes fonts i convertir-los en dades enllaçades. Tal com s'afirma a les línies estratègiques del PCC 2015-2017:¹⁵

Els mètodes existents per al control d'autoritats es basen en la construcció de punts d'accés autoritzats i únics a partir de cadenes de text o *literals* (seqüències de caràcters que configuren expressions literals o frases exactes). Aquest enfocament basat en cadenes funciona més o menys bé en un entorn tancat d'un catàleg de biblioteca tradicional, però no en un entorn obert en què les dades es comparteixen i estan enllaçades. Per aquest motiu, es necessiten identificadors únics. El web és un repte i alhora una oportunitat per a les biblioteques, perquè poden aprofitar les dades creades fora del món bibliotecari i també poden contribuir amb les dades d'autoritats perquè altres comunitats les puguin fer servir.

Un altre problema a l'hora de transformar les metadades del llegat de les biblioteques en dades enllaçades és que hi ha diverses organitzacions bibliotecàries que intenten fer-ho, amb la qual cosa hi ha diferents models conceptuals. Zepounidou, *et al.*¹⁶ comparen quatre models conceptuals, concretament els FRBR (requisits funcionals dels registres bibliogràfics), la FR-

BRoo (versió orientada a objectes dels FRBR), el BIBFRAME i l'EDM (model de dades d'Europeana). Intenten trobar les convergències entre els models i allò que tenen en comú per poder assolir-ne la interoperabilitat.

Hi ha moltes publicacions sobre dades enllaçades, però per als bibliotecaris encara és un repte entendre'n el concepte. Segons Banerjee,¹⁷ «Tot i que fa més d'una dècada que els bibliotecaris llegeixen i van a sessions sobre dades enllaçades i FRBR, continuen sense treure'n l'entrellat».

2. Tecnologia de dades enllaçades

2.1. Model de dades del marc de descripció de recursos (RDF)

Un model de dades és, en poques paraules, un resum de dades reals i les seves relacions. El model de dades més comú és el model de dades tabular, com ara un fitxer csv, que enumera les dades en un format de taula estructurat.

El model de dades de les dades enllaçades és el marc de descripció de recursos (RDF). És la manera de representar les dades o els recursos al web. L'RDF és un concepte essencial per poder comprendre les dades enllaçades, però per poder entendre'l primer hem de saber què és l'URI (identificador uniforme de recursos).

L'URI defineix l'adreça única de qualsevol cosa que hi hagi a internet, com l'adreça postal que té cada casa. Aquest «qualsevol cosa que hi hagi

14. Kimmy Szeto, «The Mystery of the Schubert song: the linked data promise» [en línia], *Notes*, v. 74, n. 1 (2017), p. 9-23, <<http://dx.doi.org/10.1353/not.2017.0071>> [Consulta: 08/02/2018].
15. Carol J. Godby, Karen Smith-Yoshimura, «From records to things: managing the transition from legacy library metadata to linked data», *Bulletin of the Association for Information Science and Technology*, v. 43, n. 2 (2017), p. 18-23, <<https://doi.org/10.1002/bul2.2017.1720430209>> [Consulta: 08/02/2018].
16. Sofia Zepounidou, Michalis Sfakakis, Christos Papatheodorou, «Representing and integrating bibliographic information into the semantic web: a comparison of four conceptual models», *Journal of Information Science*, v. 43, n. 4 (2017), p. 525-553.
17. Jim Banerjee, «Translating technobabble: all you really need to know about URIs, linked data, and FRBR», *Computers in Libraries*, v. 37, n. 10 (2017), p. 21-24.

a internet» no només inclou les entitats físiques, com ara una poma o Da Vinci, sinó que també inclou conceptes abstractes, com ara l'amor i la pau. Per exemple, l'URI (*example.org/yongming-wang*) és únic a tot internet i fa referència a un dels autors d'aquest article, Yongming Wang (cal dir que *example.org* no és un nom de domini real; és una convenció web utilitzada per mostrar un exemple d'un lloc web que qualsevol pot fer servir com a exemple). Aquest URI també és un URL. És a dir, l'URL és un tipus d'URI al web. Recentment, el nom d'URI s'ha canviat a IRI, que és l'abreviació d'International Resource Identifier (identificador internacional de recursos).

Així doncs, què és exactament l'RDF? I quina funció té en la tecnologia de dades enllaçades? L'RDF és un model de dades que s'utilitza per identificar o descriure coses (també anomenades *entitats*) i les relacions que tenen al web. Una sentència RDF, també anomenada *terna RDF* (*triple RDF*), té tres parts: subjecte, predicat i objecte. El subjecte i l'objecte designen dues coses diferents, i el predicat descriu la relació entre el subjecte i l'objecte. El format que té és aquest:

<subject> <predicate> <object>

Aquest n'és un exemple: <Bob> <is a friend of> <Alice>. Es pot expressar en un gràfic com el de la figura 1.

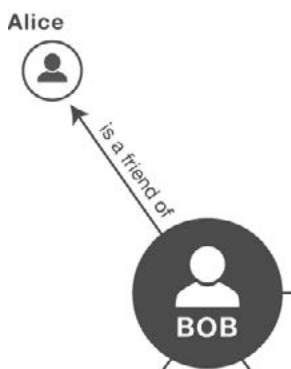


Figura 1. Part de la figura informal de les ternes d'exemple. Reimprès del text W3C RDF 1.1. Recuperat de (www.w3.org/TR/rdf11-primer/#fig1). Copyright © [24/06/2014] World Wide Web Consortium, (MIT, ERCIM, Keio, Beihang) <<http://www.w3.org/Consortium/Legal/2015/doc-license>>.

Segons els principis de les dades enllaçades, un subjecte ha d'incloure l'identificador uniforme de recursos (URI) i un objecte pot ser o bé un URI o bé una cadena de text. El predicat s'hauria de definir amb vocabulari RDF i ha de ser un URI. El vocabulari RDF és com un esquema en una base de dades relacional. La diferència és que un esquema serveix per definir l'abast i el tipus de dades, mentre que el vocabulari RDF serveix per definir la relació entre els elements de les dades. Per tant, en l'exemple anterior, el subjecte <Bob> es presenta en el format (*http://example.org/bob/*), que és un URI únic al web o un URL (HTTP URI) únic. L'objecte <Alice> és un altre URL únic, (*http://example.org/alice*). Finalment, el predicat <is a friend of> també es presenta en format URI, (*http://example.org/friendOf/*). La sentència «<Bob> <is a friend of> <Alice>» escrita com a sentència oficial RDF seria la següent:

<*http://example.org/bob*> <*http://example.org/friendOf*> <*http://example.org/alice*>

Un gràfic de ternes múltiples mostrarà elements múltiples i les seves relacions. A més, la gràcia de les dades enllaçades rau en el fet que aquests elements múltiples i les seves relacions tenen diferents localitzacions a internet (estan allotjats a diferents servidors web). Els programadors poden crear aplicacions per agregar aquests elements (dades) sobre la marxa. Els usuaris poden seguir aquests enllaços (les relacions) per trobar més informació del seu interès. La figura 2 n'és un exemple.

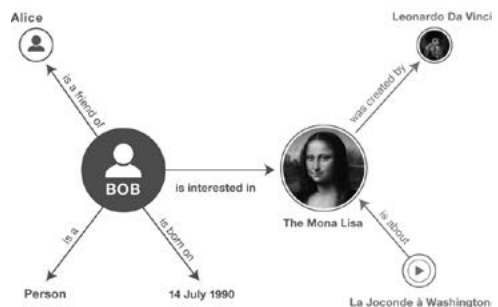
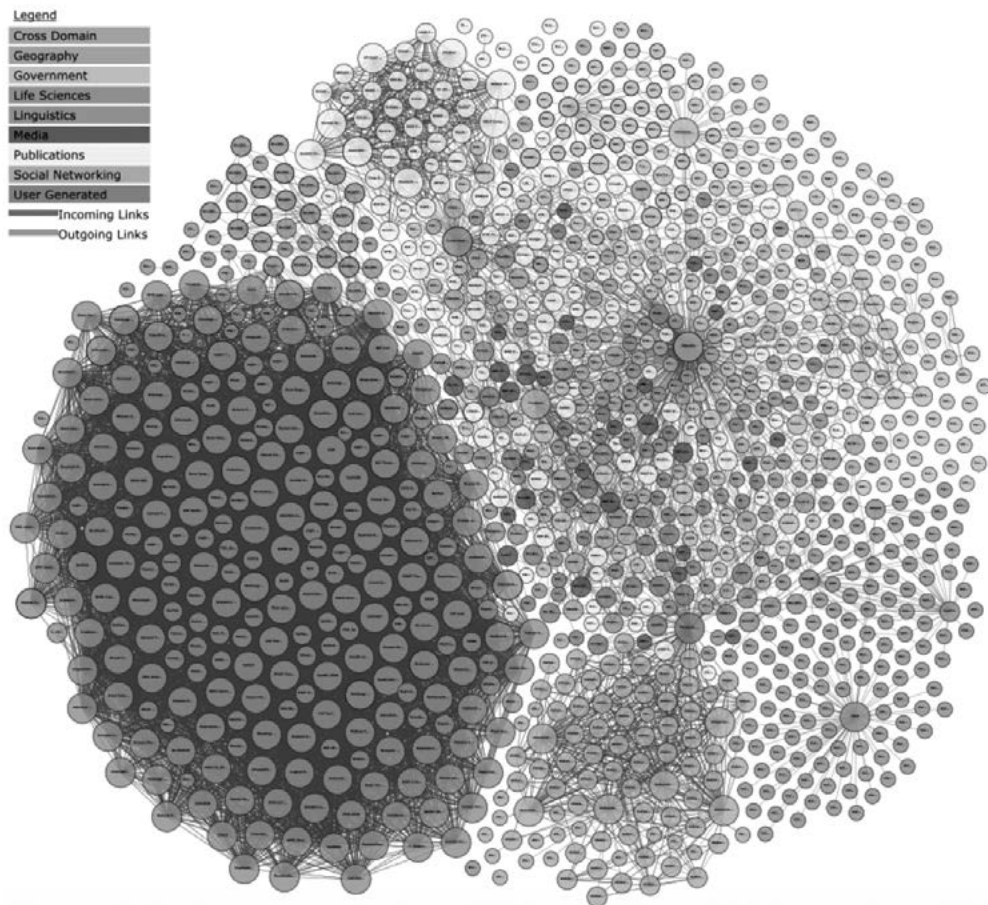


Figura 2. Figura informal de les ternes d'exemple. Reimprès del text W3C RDF 1.1. Recuperat de (www.w3.org/TR/rdf11-primer/#fig1). Copyright © [24/06/2014] World Wide Web Consortium, (MIT, ERCIM, Keio, Beihang) <<http://www.w3.org/Consortium/Legal/2015/doc-license>>.



A la figura 2 hi ha sis ternes que tenen la forma <subject> <predicate> <object>. Són:

<Bob> <is a> <person>.

<Bob> <is a friend of> <Alice>.

<Bob> <is born on> <the 4th of July 1990>.

<Bob> <is interested in> <the Mona Lisa>.

<the Mona Lisa> <was created by> <Leonardo da Vinci>.

<the video 'La Joconde à Washington'> <is about> <the Mona Lisa>.

Aquesta mena de gràfic es pot ampliar il·limitadament. Així doncs, es pot descriure de manera única gairebé tot el que hi ha al web i també es pot enllaçar. Si l'anéssim ampliant, podríem obtenir el gràfic gegant de la Figura 3.

Figura 3. Diagrama de núvol de dades obertes enllaçades 2017. Reimprès del Diagrama de núvol de dades obertes enllaçades d'Andrejs Abele, John P. McCrae, Paul Buitelaar, Anja Jentzsch i Richard Cyganiak. Recuperat de (*lod-cloud.net*). Llicència CC-BY-SA.

2.2. Vocabulari RDF

El vocabulari RDF és molt més difícil d'explicar que altres conceptes relacionats amb les dades enllaçades, com ara el model de dades RDF, del qual ja s'ha parlat, i la serialització RDF que es tractarà més endavant. La complexitat del vocabulari RDF fa que hi hagi molta confusió i s'estiguin adaptant les dades enllaçades molt lentament. En poques paraules, els vocabularis RDF són com l'esquema en una base de dades relacional. Segons Wood:¹⁸

18. David Wood, *et al.*, *op. cit.*, p. 38.



Proporcionen definicions dels termes utilitzats per establir relacions entre les dades. Tanmateix, a diferència de l'esquema de la base de dades relacional, els vocabularis RDF es distribueixen pel web i els desenvolupen persones d'arreu del món, i només s'utilitzen com a dades enllaçades si molta gent decideix fer-los servir.

El vocabulari RDF té format HTTP URI i es defineix i es guarda en llocs web coneguts. A continuació, en veurem alguns exemples. El vocabulari RDF s'utilitza en la posició de predicat de les ternes RDF per definir la relació entre un subjecte i un objecte. Amb l'exemple anterior, «Bob is a friend of Alice», mostrarem com es fa servir el vocabulari RDF.

Com s'ha vist abans, la sentència RDF per a «Bob is a friend of Alice» s'escriu així:

```
<http://example.org/bob> <http://example.org/friendOf> <http://example.org/alice>
```

Per fer servir el vocabulari RDF en la posició de predicat a la realitat, hem de reemplaçar `<http://example.org/friendOf>` per: `http://xmlns.com/foaf/0.1/knows`.

Analitzem-ho detingudament. Hi ha dues parts a `http://xmlns.com/foaf/0.1/knows`: La primera, `http://xmlns.com/foaf/0.1/`, és l'espai de noms per al vocabulari FOAF (amic d'un amic), i la segona part, `knows`, és atribut del vocabulari FOAF. FOAF és un projecte de codi obert desenvolupat a mitjans de l'any 2000 per enllaçar gent al web. FOAF s'utilitza a les xarxes socials, en molts projectes de dades enllaçades. Quan volem descriure que algú és un amic d'algú altre podem utilitzar aquest atribut FOAF i qualsevol programa informàtic de dades enllaçades l'entendrà i en reconeixerà automàticament el significat.

Com s'ha dit anteriorment, el vocabulari FOAF té el format HTTP URI (`http://xmlns.com/foaf/0.1/`), i està guardat en un lloc web ben conegut (`http://xmlns.com`).

Un espai de noms (namespace) és un marcadore de desambiguació que permet identificar de manera unívoca els conjunts de noms o atributs. A l'espai de noms de «foaf», abreviació d'`http://xmlns.com/foaf/0.1/`, tots els seus atributs, inclòs el que fem servir aquí, «knows»,

Un model de dades és, en poques paraules, un resum de dades reals i les seves relacions. El model de dades de les dades enllaçades és el marc de descripció de recursos (RDF). És la manera de representar les dades o els recursos al web.

estan guardats i definits. És a dir, no hi ha cap mena d'ambigüitat: «foaf:knows», abreviació d'<http://xmlns.com/foaf/0.1/knows>, fa referència a una relació entre dues persones. No ho podem confondre mai amb altres «knows». En aquest sentit, l'espai de noms també es pot anomenar «el prefix».

Per poder entendre bé els vocabularis RDF i, sobretot, per poder crear un vocabulari RDF, és essencial comprendre'n els dos components clau: l'esquema del marc de descripció de recursos (RDFS) i el llenguatge d'ontologies web (OWL). L'RDFS és el llenguatge de definició per al vocabulari RDF. L'RDFS defineix les classes i categories, la qual cosa ajuda a crear nous vocabularis RDF. L'OWL és una extensió de l'RDF. Els autors han optat per no aprofundir en aquests conceptes perquè són molt complexos i l'extensió de l'article no ho permet.

Un altre vocabulari RDF important és el sistema senzill d'organització del coneixement (SKOS). El principal objectiu de SKOS és convertir en vocabularis RDF els vocabularis controlats tradicionals, com ara els tesaurus i tota mena de llistes d'encapçalaments de matèria (per exemple, els Library of Congress Subject Headings). Aquesta característica fa que l'SKOS sigui especialment rellevant per a la comunitat bibliotecària.

Per poder entendre millor l'RDFS, l'OWL i l'SKOS, els autors recomanen un llibre anome-

nat *Semantic Web for the Working Ontologist*, de Dean Allemang i James Hendler.¹⁹ El llibre inclou molts exemples i està escrit de manera plana i entenedora, amb la qual cosa facilita la comprensió de nocions relativament complexes, com ara la modelització semàntica i l'ontologia.

2.3. Serialització RDF

La serialització RDF és la manera com s'escriuen les sentències RDF perquè el programa informàtic les pugui processar i llegir. Hi ha diferents tipus de serialització RDF. Les comunes són: Turtle (abreviació de *Terse RDF Triple Language*, llenguatge concís de ternes RDF), RDF/XML (el format RDF original en XML), RDFa (RDF integrat en atributs HTML), i el més nou i conegut, JSON-LD. En aquest article ens centrarem en JSON-LD.

JSON-LD, abreviació de JavaScript Object Notation (notació d'objecte de JavaScript) per enllaçar dades, es va fer conegut perquè és un dels llenguatges de programació preferits de molts desenvolupadors de webs i gairebé tots els llenguatges de programació tenen múltiples biblioteques per analitzar-lo. JSON és fàcil d'escriure i de llegir. Continuem amb l'exemple anterior per mostrar com es poden escriure les ternes RDF en format JSON-LD.

Mirem les tres sentències RDF següents:

```
<Bob> <is a> <person>.
<Bob> <is a friend of> <Alice>.
<Bob> <is born on> <the 4th of July 1990>.
```

Les seves ternes RDF formals són:

```
<http://example.org/bob> <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> <http://xmlns.com/foaf/0.1/Person>.
<http://example.org/bob> <http://xmlns.com/foaf/0.1/knows> <http://example.org/alice>.
<http://example.org/bob> <http://schema.org/birthDate> "1990-07-04"^^<http://www.w3.org/2001/XMLSchema#date>.
```

19. Dean Allemang, Jim Hendler, *Semantic web for the working ontologist: effective modeling in RDFS and OWL*, Waltham: Morgan Kaufmann; Oxford: Elsevier, 2011.

A l'exemple anterior, <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> es pot escurçar com a `rdf:type`, que pertany a l'RDFS.

<http://schema.org/birthDate> prové d'un altre vocabulari RDF conegut, `schema.org`. I `1990-07-04` és un objecte literal del tipus `data`, tal com es defineix a l'esquema XML.

El format JSON-LD és el següent:

```
{
  "@context": {
    "foaf": "http://xmlns.com/foaf/0.1/",
    "Person": "foaf:Person",
    "knows": {
      "@id": "foaf:knows",
      "@type": "@id"
    },
    "birthdate": {
      "@id": "http://schema.org/birthDate",
      "@type": "http://www.w3.org/2001/XMLSchema#date"
    }
  },
  "@id": "http://example.org/bob#me/",
  "@type": "Person",
  "birthdate": "1990-07-04",
  "knows": "http://example.org/alice#me/"
}
```

Com s'observa, el format JSON-LD és fàcil d'entendre. Tot està en parells atribut-valor (*key-value pairs*). L'única part complicada és l'objecte del context dins del qual es defineixen els espais de noms o els prefixos.

2.4. SPARQL – El llenguatge de cerca

SPARQL no és un acrònim, el nom complet és protocol SPARQL i llenguatge de cerca RDF. SPARQL és el llenguatge de cerca per al conjunt de dades RDF, de la mateixa manera que SQL és el llenguatge de cerca per a les bases de dades relacionals. Les sintaxis de SPARQL i SQL són similars, però no tenen res més en comú. De fet, per aprendre ràpidament SPAR-

QL caldria oblidar el que s'hagi après de SQL. Podem fer servir SPARQL per a cercar un fitxer RDF local amb dades RDF en forma de ternes (vegeu els exemples de més endavant). També podem utilitzar SPARQL per a fer una cerca al magatzem de dades RDF (*RDF data store*) remot, independentment del lloc de la xarxa on es trobi, sempre que aquest magatzem proporcionï un servei SPARQL de punt final (*SPARQL endpoint*). A més a més, podem combinar tant consultes remotes i locals com vulguem per aconseguir les dades que volem a la nostra aplicació. Aquesta és la veritable fortalesa de SPARQL i les dades enllaçades.

Primer, començarem amb un exemple SPARQL senzill. Demostrarem com s'ha d'utilitzar SPARQL per cercar un fitxer RDF local. Imaginem-nos que tenim un fitxer anomenat *bob.rdf* amb el contingut següent:

```
(bob.rdf)
prefix foaf: http://xmlns.com/foaf/0.1/.
prefix rdf: http://www.w3.org/1999/02/22-rdf-syntax-ns.
prefix schema: http://schema.org/.
prefix xsd: http://www.w3.org/2001/XMLSchema#.
```

```
http://example.org/bob#me rdf:type foaf:person.
http://example.org/bob#me schema:birthdate
"1990-07-04"^^xsd:date.
http://example.org/bob#me foaf:knows http://
example.org/alice#me.
http://example.org/bob#me foaf:knows http://
example.org/lisa#me.
```

Volem fer servir SPARQL per trobar tots els amics d'en Bob al fitxer *bob.rdf*. Aquesta és la cerca:

(Nota: SPARQL troba el resultat comprovant patrons. Totes les paraules amb interrogant són una variable).

```
prefix foaf: http://xmlns.com/foaf/0.1/.

select ?x
from bob.rdf
```

```
where {
  http://example.org/bob#me
  foaf:knows ?y ;
}
```

El resultat serà aquest:

X
Alice
Lisa

Dins de la clàusula en què *?y* és la variable, *http://example.org/bob#me* i *foaf:knows* són els valors donats. Demana que es busqui el valor de la posició de l'objecte a totes les ternes tenint en compte el subjecte de *http://example.org/bob#me* i el predicat de *foaf:knows*. És a dir, Alice i Lisa. Si volem trobar tant el subjecte com l'objecte amb el valor del predicat de *foaf:knows*, farem servir la cerca següent:

prefix foaf: <http://xmlns.com/foaf/0.1/>.

```
select ?x (as Name) ?y (as Friend)
from bob.rdf
where {
  ?a foaf:knows ?b ;
}
```

El resultat:

Name	Friend
Bob	Alice
Bob	Lisa

Un cop més, SPARQL troba el resultat emparellant els valors donats a la clàusula i recupera tots els valors de les variables en una o més posicions de la terna.

Aquesta és només una petita introducció a SPARQL. Es pot complicar força més quan s'aplica de debò.

3. Implicació de la comunitat bibliotecària

L'any 2005 les biblioteques van entendre el valor que tenien les dades enllaçades i el web semàntic per descriure els recursos bibliotecaris, i als EUA, Canadà i Regne Unit es va crear un comitè conjunt per revisar les regles de catalogació AACR2. El llançament de l'R-DA (recursos, descripció i accés) l'any 2010 va proporcionar les directrius per catalogar i descriure els recursos bibliotecaris, de manera que les dades bibliogràfiques resultants estiguin en consonància amb les dades enllaçades, un estàndard web reconegut i compartit per altres comunitats d'internet. Les dades enllaçades tenen múltiples avantatges, com ara la divulgació de les dades bibliogràfiques dels repositoris al web, l'enllaç a recursos d'altres comunitats i la recuperació dels recursos bibliotecaris que fan els cercadors d'internet. Segons els estudis, un 82 % dels consumidors d'informació comencen a cercar des d'un cercador d'internet i només un 1 % des d'un lloc web d'una biblioteca.²⁰ Així doncs, amb l'exposició de les dades bibliotecàries al web semàntic i a internet, els recursos bibliotecaris s'utilitzaran més i s'oferiran més bons serveis als usuaris.

D'ençà de 2010 les biblioteques empaiten amb fermesa l'objectiu de transformar les dades bibliogràfiques en dades enllaçades, que és el format necessari per al web semàntic. El camí és tortuós i l'avançament és més lent del que s'esperava. Un dels motius és que les dades enllaçades són un concepte molt nou per a les biblioteques i es tracta d'un canvi molt dràstic respecte a la pràctica de catalogació tradicional. S'han de resoldre molts tecnicismes abans de posar en marxa la nova pràctica. La manca de participació podria ser un altre contratemps. Fins ara, només les grans biblioteques i organitzacions tenen els coneixements

20. Online Computer Library Center, «Meeting the e-resources challenge: an OCLC report on effective management, access and delivery of electronic collections» [en línia], [Dublin, EUA: OCLC, 2013], <<https://www.oclc.org/content/dam/oclc/reports/pdfs/OCLC-E-Resources-Report-US.pdf>> [Consulta: 08/02/2018]; «Google Ads: what are Google Ads & how do they work?» [en línia], *Wordstream.com* [Boston: Wordstream, cop. 2018], <<http://www.wordstream.com/google-ads>> [Consulta: 08/02/2018].

tècnics i els recursos financers necessaris per provar i desenvolupar projectes de dades enllaçades. La LC, l'OCLC i altres biblioteques nacionals estan al capdavant dels projectes de dades enllaçades a les biblioteques. La majoria de petites biblioteques observen i esperen, però no hi participen gaire. Manquen prototips que demostrin els beneficis de tenir les dades bibliotecàries com a dades enllaçades i molts bibliotecaris encara no s'imaginen com serà i com funcionarà el futur catàleg semàntic. No és gens senzill transformar la ingent quantitat de dades catalogades durant quaranta anys en dades enllaçades. Les ontologies i els vocabularis de les biblioteques són complexos i es necessita molt de temps per completar-los. Tanmateix, tot i aquests obstacles, les biblioteques d'arreu del món han apostat per les dades enllaçades, i s'ha progressat molt. A continuació, es descriuen els assoliments de la comunitat bibliotecària en el camp de les dades enllaçades i del web semàntic.

3.1. Library of Congress (LC)

La LC ha liderat mundialment la promoció de les tecnologies de dades enllaçades i les seves possibles aplicacions a les biblioteques. El primer que va fer la Library of Congress va ser convertir els Library of Congress Subject Headings (LSCHE), els registres d'autoritat de noms propis (Name Authority Files) i altres llenguatges controlats en sentències RDF i URI, i va aconseguir així que altres aplicacions del web semàntic els poguessin utilitzar. La LC també té un paper fonamental en el desenvolupament de les regles de catalogació RDA, que es basen en els FRBR (requisits funcionals dels registres bibliogràfics) i donen suport a les dades enllaçades. Després del llançament de l'RDA l'any 2011, la LC va començar a treballar en el BIBFRAME, que és un nou estàndard de visualització que pretén reemplaçar els estàndards

No és gens senzill transformar la ingent quantitat de dades catalogades durant quaranta anys en dades enllaçades. Les ontologies i els vocabularis de les biblioteques són complexos i es necessita molt de temps per completar-los.

MARC. A finals del 2012 i principis del 2013, es va llançar el BIBFRAME 1.0 com a prova en un projecte pilot. Incloua un conjunt d'eines, com ara el BIBFRAME Editor, el MARC to BIBFRAME Comparison Viewer i la MARCXML to BIBFRAME Transformation Tool. Des d'aleshores, la LC ha anat provant i modificant el BIBFRAME. Es tracta d'un procés complicat i llarg. Els vocabularis o ontologies BIBFRAME són la base i també la part més difícil del desenvolupament del BIBFRAME. La conversió de l'actual MARC al BIBFRAME és un altre repte. La LC té dinou milions de registres MARC.²¹

L'últim model de dades, la segona generació del BIBFRAME, és el BIBFRAME 2.0, llançat l'any 2016. El nou model de dades revisat inclou tres categories d'abstracció: obra, manifestació i exemplar, i defineix tres conceptes més relacionats amb aquestes categories: agents, matèries i esdeveniments.²² El BIBFRAME 2.0 ha llançat programes i requisits de conversa per passar dels estàndards MARC al BIBFRAME. Tanmateix, el BIBFRAME Editor 2.0 encara s'està desenvolupant, i inclourà ontologies més completes amb classes i atributs especialment dissenyats per descriure recursos bibliotecaris. Tots dos vocabularis o ontologies principals utilitzats al BIBFRAME 2.0 són el BIBFRAME i el MADS/RDF (Metadata Authority Description Schema in RDF). A més a més, el BIBFRAME 2.0 es basa en algunes

21. Sally McCallum, «BIBFRAME development» [en línia], *JLIS.it: Italian Journal of Library, Archives, and Information Science*, v. 8, n. 3 (2017), p. 71-85, <<https://www.jlis.it/article/view/12415>> [Consulta: 27/03/2018].

22. «Overview of the BIBFRAME 2.0 Model» [en línia], [Washington: Library of Congress, April 21, 2016], <<https://www.loc.gov/bibframe/docs/bibframe2-model.html>> [Consulta: 27/03/2018].



ontologies bàsiques desenvolupades pel World Wide Web Consortium, l'OWL (llenguatge d'ontologies web), l'RDFS (marc de descripció de recursos) i l'SKOS (sistema senzill d'organització del coneixement). «L'ontologia BIBFRAME 2 està molt més ben integrada a l'entorn RDF i també molt més sincronitzada amb les regles de catalogació RDA, malgrat no seguir cap regla concreta».²³

Tot i la complexitat de les ontologies i els vocabularis, el BIBFRAME Editor en si és una eina senzilla que convertirà les dades bibliogràfiques a través d'una interfície d'entrada basada en el web en sentències RDF, un dels components bàsics de les dades enllaçades. S'espera que els programes de conversió BIBFRAME 2.0 puguin processar encara més registres MARC i incloure'n més dades. Encara no se sap com i on se cercaran i mostraran les dades BIBFRAME.

La LC ha avançat moltíssim en el desenvolupament del BIBFRAME. No cal dir que el BIBFRAME és un projecte en curs que tindrà ac-

tualitzacions dels vocabularis i llançaments de noves versions molt després d'estar operatiu.

3.2. Online Computer Library Center (OCLC)

L'OCLC és una altra entitat líder en projectes i recerca de dades enllaçades per a les biblioteques. La majoria de projectes de dades enllaçades de l'OCLC estan relacionats amb WorldCat.org, una base de dades de més de 400 milions de registres bibliogràfics de més de 16.000 biblioteques.²⁴ Amb la col·laboració de la LC i altres biblioteques nacionals, l'OCLC ha aconseguit grans fites en aquest sentit.

El primer projecte públic i notori que va dur a terme l'OCLC va ser afegir el marcatge Schema.org als seus registres del WorldCat. Schema.org va ser creat pels principals cercadors d'internet, com ara Google, Bing i Yandex, i proporciona especificacions i requisits combinats perquè qualsevol persona o organització els faci servir si volen que els cerquin i visua-

23. Sally McCallum, *op. cit.*

24. Online Computer Library Center, «OCLC and linked data» [en línia], [Dublin, EUA: OCLC Linked Data Research, 2017], <https://www.oclc.org/content/dam/oclc/services/brochures/215912_WWAE-OCLC-Linked-Data-Report.pdf> [Consulta: 08/02/2018].

litzin com a dades enllaçades. «En afegir el marcatge Schema.org a tots els llibres, revistes i altres recursos bibliogràfics a WorldCat.org, tota la versió disponible de WorldCat ara també la poden fer servir les aplicacions intel·ligents rastrejadores del web (*crawlers*), com ara les dels cercadors Google i Bing, que poden utilitzar aquestes metadades en índexs de cerca i altres aplicacions».²⁵ Com que el vocabulari d'Schema.org és més general i no és prou detallat per descriure els recursos bibliotecaris, l'OCLC va liderar i participar en el procés per reconciliar el vocabulari d'Schema.org amb el vocabulari de BIBFRAME, així com el desenvolupament de l'extensió bibliogràfica del vocabulari de Schema.org (*bib.schema.org*).

L'OCLC va implementar WorldCat.org Works, perquè totes les manifestacions de la mateixa obra estiguin enllaçades i mostrades en un clúster que fa servir l'algorisme de conjunt d'obres FRBR. «L'algorisme recull els registres bibliogràfics i els agrupa segons la informació del títol i de l'autor dels registres d'autoritat i bibliogràfics».²⁶ Els estàndards dels cercadors d'internet se segueixen perquè «l'entitat WorldCat Work es basa en les propietats definides per la categoria schema: CreativeWork».²⁷ L'avantatge de reunir tots els formats d'una obra a partir del títol és evident. D'ençà de juliol de 2017, uns 215 milions d'entitats d'obres estan disponibles a WorldCat.org.²⁸

OCLC Persons és un projecte semblant, però vinculat amb entitats personals. «Les entitats personals de WorldCat relacionen la informació d'una persona amb una breu descripció que inclou variants del seu nom, les seves obres i les seves fonts biogràfiques. Des de juliol de 2017, les entitats personals de WorldCat inclou-

uen més de 117 milions de descripcions d'autors, directors, músics i altres, que s'han extret directament de WorldCat. Aquestes entitats es van fer servir en un programa pilot de dades enllaçades en què les biblioteques van utilitzar les persones de WorldCat en el seu flux de treball regular».²⁹

El registre d'autoritat virtual internacional (VIAF) és un altre projecte de dades enllaçades exitós promogut per l'OCLC i diverses biblioteques nacionals, incloses la LC, la Deutsche National Bibliothek i la Bibliothèque Nationale de France. VIAF (*viaf.org*) és un registre d'autoritat internacional que es basa en dades d'autoritats d'una llista de biblioteques nacionals i que manté l'OCLC. La funció que té és la següent: «VIAF emparella i enllaça els registres d'autoritat de les biblioteques nacionals. Després agrupa tots els registres d'autoritat d'una entitat concreta en un «súper» registre d'autoritat fusionat que agrupa les diferents descripcions d'aquella entitat».³⁰ VIAF API permet que els usuaris cerquin dades d'autoritats amb paraules clau, nom, títol i més paràmetres, i recuperin registres d'autoritat i relacions entre aquests registres. VIAF funciona amb la llicència Open Data Commons Attributions (ODC-by) i qualsevol persona o organització la pot fer servir. «Des de l'any 2009 VIAF està disponible com a dades enllaçades i ara mateix és un dels recursos de dades enllaçades més utilitzat publicat per la comunitat bibliotecària».³¹

L'OCLC i la LC han desenvolupat conjuntament FAST (Faceted Application of Subject Terminology), un esquema de llistes d'encapçalaments de matèria general que prové dels LCSH. L'objectiu de FAST és crear un altre esquema de matèries amb múltiples facetes

25. Bob Murphy, «OCLC adds linked data to WorldCat.org» [en línia], *OCLC.org: news releases* (June 20, 2012), <<http://worldcat.org/arcviewer/7/OCC/2015/03/19/H1426803137790/viewer/file1303.html>> [Consulta: 08/02/2018].

26. Online Computer Library Center, «OCLC and linked data», *op. cit.*

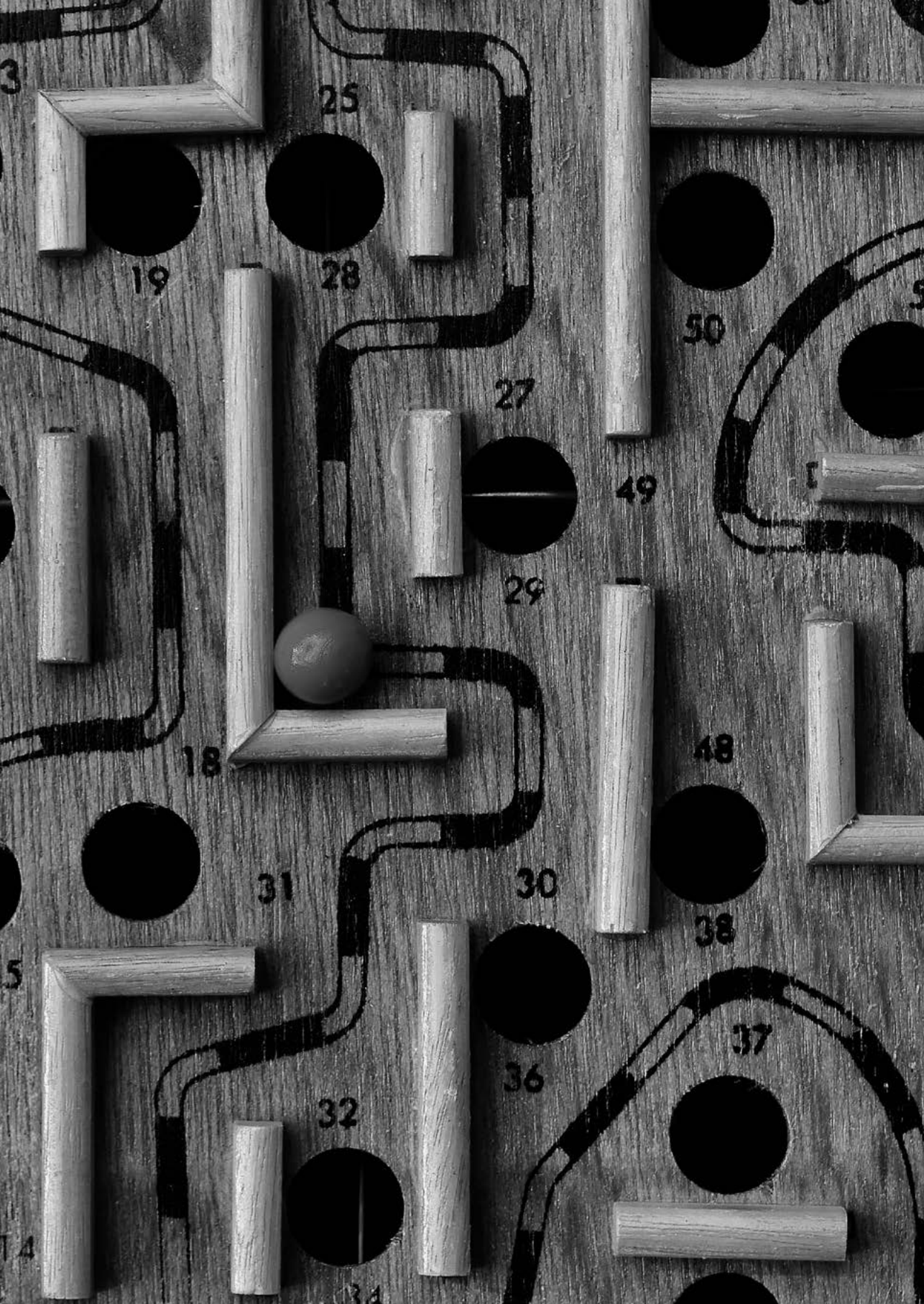
27. Online Computer Library Center, «WorldCat work descriptions» [en línia], [Dublin, EUA: OCLC Developer Network, 2017], <<https://www.oclc.org/developer/develop/linked-data/worldcat-entities/worldcat-work-entity.en.html>> [Consulta: 08/02/2018].

28. Online Computer Library Center, «OCLC and linked data», *op. cit.*

29. Online Computer Library Center, «OCLC and linked data», *op. cit.*

30. Online Computer Library Center, «OCLC and linked data», *op. cit.*

31. Online Computer Library Center, «OCLC and linked data», *op. cit.*



que sigui fàcil d'usar i senzill d'entendre. Les dues llistes d'encapçalaments de matèria són compatibles i els LCHS es poden convertir en FAST. D'ençà de 2011, FAST està disponible com a dades enllaçades. Algunes organitzacions i biblioteques nacionals utilitzen FAST per a la descripció de metadades i la indexació de matèries. Segons l'OCLC, FAST és «un dels vocabularis de matèries més utilitzat de l'àmbit bibliotecari».³²

Gràcies als grans projectes de dades enllaçades com ara WorldCat Works, les entitats de persones de WorldCat i el marcatge de Schema.org, «es van fer més de 74 milions de visites a WorldCat.org l'any 2016 i més de disset milions de visites a catàlegs de biblioteques locals de tot el món».³³

3.3. Altres projectes de dades enllaçades de biblioteques dels Estats Units (EUA)

El moviment bibliotecari de dades enllaçades també inclou els projectes de Zepheira i moltes altres biblioteques acadèmiques. Eric Miller, director general de Zepheira, una empresa consultora de dades enllaçades que va desenvolupar BIBFRAME 1.0, defensa l'acció immediata de les biblioteques perquè publiquin les seves dades al web, de manera que els cercadors d'internet les podran buscar i mostrar a dalt de tot de la pàgina de resultats. Per aconseguir aquest objectiu, Zepheira va promoure la iniciativa Libhub l'any 2014 i Library.Link Network l'any 2016. Associat amb proveïdors com ara EBSCO, SirsiDynix i Innovative Interfaces, el projecte Library.Link Network implica un procés de quatre passos en el qual «Zepheira copia el catàleg d'una biblioteca associada, converteix els registres en el format estructurat BIBFRA-

ME i allotja aquests registres BIBFRAME a la xarxa mundial de distribució de contingut compartit Library.Link, dissenyada per a la ingesta a gran escala» i, a més, «s'afegeix a cada registre la llicència Creative Commons, que ha d'estar assignada a la biblioteca perquè els proveïdors de serveis com ara Google i Microsoft sàpiguen d'on provenen les dades i què poden fer-ne les empreses».³⁴ El pas final és publicar dades bibliogràfiques, esdeveniments, horaris i informació del personal de les biblioteques al web. Els participants principals són les biblioteques públiques i per ara el projecte avança.

La conversió de dades és un component clau per desenvolupar dades enllaçades per a biblioteques. El Colorado College porta a terme dos projectes: «Un per convertir no només els estàndards MARC sinó també altres dades que tenen en formats com ara MODS, Dublin Core i altres fitxers XML al BIBFRAME RDF, per poder accedir a tots aquests fitxers. Un altre converteix els registres dels estàndards MARC al BIBFRAME i després converteix el BIBFRAME a Schema.org per enviar-los a Google».³⁵

Unes quantes biblioteques acadèmiques grans van rebre ajuts de l'Andrew W. Mellon Foundation per col·laborar en projectes de dades enllaçades de 2014 a 2018.³⁶ Les universitats que participen en el projecte són Cornell, Harvard, Columbia, Stanford i Princeton, entre d'altres. Els projectes que han rebut la subvenció són Linked Data for Libraries (LD4L), Linked Data for Libraries Labs (LD4L Labs) i Linked Data for Production (LD4P). Els objectius d'aquests projectes són crear una ontologia compatible amb el BIBFRAME i altres ontologies existents per descriure col·leccions acadèmiques locals, desenvolupar un sistema semàntic de font oberta

32. Online Computer Library Center, «OCLC and linked data», *op. cit.*

33. Online Computer Library Center, «OCLC and linked data», *op. cit.*

34. Matt Enis, «Library.Link builds open web visibility for library catalogs, events» [en línia], *Library Journal: news* (Jun 22, 2016), <<https://www.libraryjournal.com/?detailStory=library-link-builds-open-web-visibility-for-library-catalogs-events>> [Consulta: 27/03/2018].

35. Sally McCallum, *op. cit.*

36. «LD4L: linked data for libraries» [en línia], [Ithaca: LD4P, cop. 2014-2018] <<https://www.ld4l.org/>> [Consulta: 27/03/2018].



per editar, cercar i mostrar els recursos acadèmics, comprovar el flux de treball en els serveis tècnics de dades enllaçades i crear eines i directrius per al futur. La University of California Davis Library també va comprovar el projecte BIBFLOW per estudiar-ne el flux de treball. Tots aquests esforços contribueixen a la tasca de la LC en matèria de dades enllaçades i totes les biblioteques se'n beneficiaran.

La National Library of Medicine també va participar activament en el desenvolupament de l'ontologia i la prova del BIBFRAME. L'any 2014 la NLM va publicar versions beta de dos conjunts de dades com a dades enllaçades: PubChemRDF, amb informació de les activitats biològiques de molècules petites, i MeSH RDF, el tesaurus de la NLM de les llistes d'encapçalaments de matèria mèdica. Tots dos productes RDF es poden cercar des de les seves pròpies interfícies de cerca SPARQL, o la cerca es pot integrar directament en programes i serveis que utilitzin els punts finals SPARQL.³⁷

3.4. Proveïdors de sistemes de gestió de biblioteques

El BIBFRAME Editor 2.0 encara no s'ha llançat i això dificulta que un proveïdor de sistemes de gestió de biblioteques inverteixi diners i personal en un model de dades que ara mateix encara s'està desenvolupant. Tanmateix, alguns proveïdors han expressat el compromís d'utilitzar les dades enllaçades i la intenció d'incorporar el BIBFRAME als seus sistemes. Uns quants ja s'estan preparant per al llançament del BIBFRAME Editor 2.0.

Ex Libris està desenvolupant el mòdul de publicació BIBFRAME, que transformarà les dades MARC en dades BIBFRAME. El full de ruta de l'empresa inclou la catalogació en format BIBFRAME i la descoberta dels materials catalogats en tots els formats a Primo, inclosos els de dades enllaçades. Innovative Interfaces, Inc. i SirsiDynix es van associar amb Zepheira per afegir més funcions al sistema existent, la qual cosa permetrà que les biblioteques transformin les dades MARC en dades BIBFRAME. També incorporaran les dades d'emplaçament de la biblioteca perquè els usuaris la puguin

37. «BIBFLOW. XI. Survey of current library linked data implementation» [en línia], [Davis: University of California Davis Library, 2016], <<https://bibflow.library.ucdavis.edu/xi-survey-of-current-library-linked-data-implementation/>> [Consulta: 08/02/2018].

localitzar a través de la pantalla. Milloraran les capes de descoberta (*discovery layers*) per revelar les dades enllaçades i connectar-se a recursos externs per obtenir contingut enriquit.

El setembre de 2017 els bibliotecaris de setze països europeus i els Estats Units es van reunir a Alemanya i van parlar dels impediments a l'hora d'implementar el BIBFRAME. En la seva opinió, la manca d'interès dels proveïdors de sistemes integrats de gestió de biblioteques (ILS) és un dels principals obstacles. El debat va originar la publicació «BIBFRAME Expectations for ILS vendors» el febrer de 2018.³⁸

3.5. Projectes de dades enllaçades en biblioteques fora dels Estats Units

Les biblioteques del món observen de prop el desenvolupament del BIBFRAME 2.0 i es preparen per al llançament del nou estàndard de visualització. Les biblioteques a Europa es van interessar per les tecnologies del web semàntic i de dades enllaçades molt abans que les biblioteques dels Estats Units. Les biblioteques europees són pioneres en tecnologies del web semàntic. El primer catàleg bibliotecari conegut que integrava dades enllaçades és LIBRIS, el catàleg del sindicat suec. L'any 2008 les dades del catàleg ja estaven disponibles com a dades enllaçades i ara conté enllaços a Wikipedia, DBpedia, registres d'autoritat de la LC (noms i matèries) i el VIAF.³⁹

Des de 2011 la British Library publica la seva bibliografia nacional britànica (Linked Open BNB) com a dades enllaçades. No disposem d'estadístiques sobre com s'està utilitzant. La Bibliothèque nationale de France (BNF) ha estat implicada en el projecte data.bnf.fr, que té l'objectiu de convertir les dades del catà-

leg de la BNF en dades enllaçades, es vol aconseguir permetre que els usuaris accedeixin a les dades bibliotecàries des del web i enllacin les dades de la BNF a DBpedia, el VIAF i altres fonts.⁴⁰ La Deutsche National Bibliothek (DNB) està desenvolupant un servei de dades enllaçades i té un compromís a llarg termini amb el web semàntic, i des de 2010 proporciona les seves dades a l'estàndard RDF. La Biblioteca Nacional Española (BNE) té un projecte semblant anomenat *datos.bne.es*, que té l'objectiu de llançar les seves dades bibliogràfiques com a dades enllaçades i, a la llarga, formar part del web semàntic.

La Canadian Linked Data Initiative (CLDI) és una col·laboració entre les cinc biblioteques de recerca més grans del Canadà, incloses la National Library of Canada (Library and Archives Canada) i la Bibliothèque et Archives Nationales du Québec. Les biblioteques participants veien que es quedaven enrere en molts sentits per al canvi imminent dels estàndards MARC a les dades enllaçades i al BIBFRAME. L'objectiu de la iniciativa és que les biblioteques canadenques estiguin al dia en la planificació estratègica per acollir-se als canvis del control bibliogràfic. Els participants treballen en la formació del personal, la preparació de les dades, el procés de descoberta millorat i tot el que cal perquè les biblioteques canadenques facin una bona transició cap al món de les dades enllaçades.

La Japanese National Library, també coneguda com a National Diet Library (NDL), proporciona metadades amb dades enllaçades obertes (LOD) per facilitar-ne l'ús per part de les aplicacions o sistemes informàtics. La National Library of China (NLC) està totalment implicada en la recerca i els debats de les tecnologies del web semàntic i dades enllaçades en llengua xinesa.

38. Organizer Group 2018 European BIBFRAME Workshop, «BIBFRAME Expectations for ILS tenders» [en línia], [Frankfurt: s.n., February 2018], <<https://wiki.dnb.de/display/EBW/Documents+and+Results>> [Consulta: 08/02/2018].

39. Ioannis Papadakis, Konstantinos Kyprianos, Michalis Stefanidakis, «Linked data URIs and libraries: the story so far» [en línia], *D-Lib Magazine*, v. 21, n. 5/6 (2015), <<https://doi.org/10.1045/may2015-papadakis>> [Consulta: 08/02/2018].

40. Ioannis Papadakis, Konstantinos Kyprianos, Michalis Stefanidakis, *op. cit*

4. Conclusions

Fa gairebé vint anys que es van crear els FRBR, després va arribar l'RDA, i ara tenim el desitjat BIBFRAME. El camí de les dades enllaçades ha estat tortuós, però no hi ha marxa enrere. El BIBFRAME s'anirà modificant, fins i tot amb el proper llançament del BIBFRAME Editor 2.0. Esperem que en els propers cinc o deu anys, la majoria de dades bibliotecàries, inclosos els milions de registres bibliogràfics que hi ha als repositoris, apareixeran com a dades enllaçades, i que es podran cercar i s'hi podrà accedir de forma gratuïta i oberta al web. Les biblioteques, no obstant això, s'hauran d'enfrontar al repte de fer aparèixer les dades bibliogràfiques a les rutes de cerca dels cercadors d'internet. «Cal que les biblioteques donin suport als nous mitjans de descoberta, i també tenen una àmplia experiència a produir dades estructurades d'alta qualitat, amb la qual cosa complementen el procés de les dades enllaçades».⁴¹ El que les biblioteques intenten assolir beneficiarà la societat i, per aquest motiu, segur que ho aconseguiran. «La comunitat bibliotecària està preparada per assolir grans avenços amb les tecnologies del web semàntic, tal com s'ha demostrat en els projectes recents del BIBFRAME, un protocol que es considera en gran part l'estàndard de la següent generació per atribuir i gestionar metadades bibliogràfiques».⁴²

Bibliografía

ALLEMANG, Dean; HENDLER, Jim. *Semantic web for the working ontologist: effective modeling in RDFS and OWL*. Waltham: Morgan Kaufmann; Oxford: Elsevier, 2011.

BANERJEE, Jim. «Translating technobabble: all you really need to know about URLs, linked data, and FRBR». *Computers in Libraries*, v. 37, n. 10 (2017), p. 21-24.

BERNERS-LEE, Tim. «Linked Data» [en línia]. [Cambridge: The World Wide Web Consortium (W3C), date: 2006/07/27, last change: 2009/06/18]. <<https://www.w3.org/DesignIssues/LinkedData.html>> [Consulta: 08/02/2018].

—; FISCHETTI, Mark. *Weaving the web: the original design and ultimate destiny of the World Wide Web*. San Francisco: Harper, 2000.

«BIBFLOW. XI. Survey of current library linked data implementation» [en línia]. [Davis: University of California Davis Library, 2016]. <<https://bibflow.library.ucdavis.edu/xi-survey-of-current-library-linked-data-implementation/>> [Consulta: 08/02/2018].

COYLE, Karen. «Linked data tools: connecting on the web» [en línia]. *Library Technology Reports*, v. 48, n. 4 (2012). <<http://dx.doi.org/10.5860/ltr.48n4>> [Consulta: 08/02/2018].

ENIS, Matt. «Library.Link builds open web visibility for library catalogs, events» [en línia]. *Library Journal: news* (Jun 22, 2016). <<https://www.libraryjournal.com/?detailStory=library-link-builds-open-web-visibility-for-library-catalogs-events>> [Consulta: 27/03/2018].

GLASER, Hugh; *et al.* «Research on linked data and co-reference resolution» [en línia]. En: International Conference on Dublin Core and Metadata Applications (2009: Seül). *DC-2009: proceedings of the International Conference on Dublin Core and Metadata Applications*. [Seoul: National Library of Korea, 2009], p. 113-117. <<http://dcpapers.dublincore.org/pubs/article/view/958/957>> [Consulta: 08/02/2018].

41. Tom Heath, Christian Bizer, *Linked data: evolving the web into a global data space*, Milton Keynes: Morgan & Claypool, cop. 2011.

42. Larry Johnson, *et al.*, *NMC Horizon Report: 2015 Library edition* [en línia], Austin: The New Media Consortium, 2015, <<https://www.nmc.org/publication/nmc-horizon-report-2015-library-edition/>> [Consulta: 08/02/2018].



Con nuestro *software* de gestión los datos catalográficos viajan, ¡por sí solos!, a los grandes agregadores y repositorios, como Hispana, Europeana o WorldCat

Digitalización enriquecida

Más información: www.digibis.com

- GODBY, Carol J.; SMITH-YOSHIMURA, Karen. «From records to things: managing the transition from legacy library metadata to linked data». *Bulletin of the Association for Information Science and Technology*, v. 43, n. 2 (2017), p. 18-23. <<https://doi.org/10.1002/bul2.2017.1720430209>> [Consulta: 08/02/2018].
- «GOOGLE Ads: what are Google Ads & how do they work?» [en línia]. *Wordstream.com* [Boston: Wordstream, cop. 2018]. <<http://www.wordstream.com/google-ads/>> [Consulta: 08/02/2018].
- HEATH, Tom; BIZER, Christian. *Linked data: evolving the web into a global data space*. Milton Keynes: Morgan & Claypool, cop. 2011.
- JIN, Qiang; HAHN, Jim; CROLL, Gretchen. «BIBFRAME transformation for enhanced discovery». *Library Resources & Technical Services*, v. 60, n. 4 (2016), p. 223-235. <<http://dx.doi.org/10.5860/lrts.60n4.223>> [Consulta: 08/02/2018].
- JOHNSON, Larry, et al. *NMC Horizon Report: 2015 Library edition* [en línia]. Austin: The New Media Consortium, 2015. <<https://www.nmc.org/publication/nmc-horizon-report-2015-library-edition/>> [Consulta: 08/02/2018].
- «LD4: linked data for libraries» [en línia]. [Ithaca: LD4P, cop. 2014-2018] <<https://www.ld4l.org/>> [Consulta: 27/03/2018].
- «LIBRARY Linked Data Incubator Group Final Report» [en línia]. [Cambridge: The World Wide Web Consortium (W3C), date: 2011/10/25]. <<https://www.w3.org/2005/Incubator/ld/XGR-ld-20111025/>> [Consulta: 27/03/2018].
- MCCALLUM, Sally. «BIBFRAME development» [en línia]. *JLIS.it: Italian Journal of Library, Archives, and Information Science*, v. 8, n. 3 (2017), p. 71-85. <<https://www.jlis.it/article/view/12415>> [Consulta: 27/03/2018].
- MILES, Alistair; et al. «SKOS core: simple knowledge organisation for the web» [en línia]. En: International Conference on Dublin Core and Metadata Applications (2005: Madrid). *DC-2005: proceedings of the International Conference on Dublin Core and Metadata Applications*. [Madrid: Universidad Carlos III, 2005], p. 3-10. <<http://dcpapers.dublincore.org/pubs/article/view/798>> [Consulta: 08/02/2018].
- MITCHELL, Erik T. «Library linked data: research and adoption» [en línia]. *Library Technology Reports*, v. 49, n. 5 (2013). <<http://dx.doi.org/10.5860/ltr.49n5>> [Consulta: 08/02/2018].
- «Library linked data: early activity and development» [en línia]. *Library Technology Reports*, v. 52, n. 1 (2016). <<http://dx.doi.org/10.5860/ltr.52n1>> [Consulta: 08/02/2018].
- MURPHY, Bob. «OCLC adds linked data to WorldCat.org» [en línia]. *OCLC.org: news releases* (June 20, 2012). <<http://worldcat.org/arcviewer/7/OCC/2015/03/19/H1426803137790/viewer/file1303.html>> [Consulta: 08/02/2018].
- ONLINE Computer Library Center. «Meeting the e-resources challenge: an OCLC report on effective management, access and delivery of electronic collections» [en línia]. [Dublin, EUA: OCLC, 2013]. <<https://www.oclc.org/content/dam/oclc/reports/pdfs/OCLC-E-Resources-Report-US.pdf>> [Consulta: 08/02/2018].
- «OCLC and linked data» [en línia]. [Dublin, EUA: OCLC Linked Data Research, 2017]. <https://www.oclc.org/content/dam/oclc/services/brochures/215912_WWAE-OCLC-Linked-Data-Report.pdf> [Consulta: 08/02/2018].
- «WorldCat work descriptions» [en línia]. [Dublin, EUA: OCLC Developer Network, 2017]. <<https://www.oclc.org/developer/develop/linked-data/worldcat-entities/worldcat-work-entity.en.html>> [Consulta: 08/02/2018].
- ORGANIZER Group 2018 European BIBFRAME Workshop. «BIBFRAME Expectations for ILS tenders» [en línia]. [Frankfurt: s.n., February 2018]. <<https://wiki.dnb.de/display/EBW/Documents+and+Results>> [Consulta: 08/02/2018].
- «OVERVIEW of the BIBFRAME 2.0 Model» [en línia]. [Washington: Library of Congress, April 21, 2016]. <<https://www.loc.gov/bibframe/docs/bibframe2-model.html>> [Consulta: 27/03/2018].
- PAPADAKIS, Ioannis; KYPRIANOS, Konstantinos; STEFANIDAKIS, Michalis. «Linked data URIs and libraries: the story so far» [en línia]. *D-Lib Magazine*, v. 21, n. 5/6 (2015). <<https://doi.org/10.1045/may2015-papadakis>> [Consulta: 08/02/2018].
- SZETO, Kimmy. «The Mystery of the Schubert song: the linked data promise» [en línia]. *Notes*, v. 74, n. 1 (2017), p. 9-23. <<http://dx.doi.org/10.1353/not.2017.0071>> [Consulta: 08/02/2018].
- THARANI, Karim. «Linked data in libraries: a case study of harvesting and sharing bibliographic metadata with BIBFRAME» [en línia]. *Information and Library Technologies*, v. 34, n. 1 (2015), p. 5-19. <<https://doi.org/10.6017/ital.v34i1.5664>> [Consulta: 08/02/2018].
- WOOD, David, et al. *Linked data: structured data on the web*. Shelter Island: Manning, 2014.
- ZAPOUNIDOU, Sofia; SFAKAKIS, Michalis; PAPAETHODOROU, Christos. «Representing and integrating bibliographic information into the semantic web: a comparison of four conceptual models». *Journal of Information Science*, v. 43, n. 4 (2017), p. 525-553. ■



Rosa FABEIRO

Unitat de Procés Tècnic, CRAI
Universitat de Barcelona
rfabeiro@ub.edu

Carme MASAGUÉ

Unitat de Procés Tècnic, CRAI
Universitat de Barcelona
cmasague@ub.edu

Anna ROVIRA

Unitat de Procés Tècnic, CRAI
Universitat de Barcelona
arovira@ub.edu

Article rebut el novembre de 2018;
revisat el desembre de 2018.

Evolució de les dades obertes enllaçades (LOD) en el context dels vocabularis controlats

Resum: A mesura que les grans biblioteques han publicat els seus conjunts de dades amb la tecnologia Linked Open Data (LOD), el nombre de serveis i projectes de dades obertes enllaçades en biblioteques no ha parat de créixer. Malgrat les dificultats inicials, l'adaptació als nous models, normatives i formats, la comunitat bibliotecària s'ha incorporat al món de les dades obertes enllaçades com a mitjà per a aconseguir un millor accés i utilització de les dades de les seves col·leccions i per a fomentar-ne la reutilització al web. Aquest article recull informació sobre els serveis i projectes de LOD actuals de biblioteques, amb especial atenció als catàlegs d'autoritats a través de la revisió de la bibliografia publicada els darrers anys. El Centre de Recursos per a l'Aprenentatge i la Investigació de la Universitat de Barcelona se suma ara a la comunitat de les LOD oferint el seu conjunt de dades d'autoritat de matèria, el Thesaurus de la Universitat de Barcelona (THUB), en dades obertes enllaçades.

Paraules clau: Dades enllaçades, vocabularis controlats, dades obertes, Thesaurus de la Universitat de Barcelona (UB).

Evolución de los datos abiertos enlazados (LOD) en el contexto de los vocabularios controlados

Resumen: A medida que las grandes bibliotecas han publicado sus conjuntos de datos con la tecnología Linked Open Data (LOD), el número de servicios y proyectos de datos abiertos enlazados en bibliotecas no ha parado de crecer. A pesar de las dificultades iniciales, la adaptación a los nuevos modelos, normativas y formatos, la comunidad bibliotecaria se ha incorporado al mundo de los datos abiertos enlazados como medio para lograr un mejor acceso y utilización de los datos de sus colecciones y para fomentar su reutilización en la web. Este artículo recoge información sobre los servicios y proyectos LOD actuales de bibliotecas con especial atención a los catálogos de autoridades a través de la revisión

de la bibliografía publicada en los últimos años. El Centre de Recursos per a l'Aprenentatge i la Investigació de la Universitat de Barcelona se suma ahora a la comunidad LOD ofreciendo su conjunto de datos de autoridad de materia, el Thesaurus de la Universitat de Barcelona (THUB) en datos abiertos enlazados.

Palabras clave: Datos enlazados, vocabularios controlados, datos abiertos, Thesaurus de la UB.

Evolution of linked open data (LOD) in the context of controlled vocabularies

Abstract: As the major libraries have gradually published their datasets with Linked Open Data (LOD) technology, the number of open data services and projects linked to libraries has continued to grow. Despite the initial difficulties, the adaptation to the new models, regulations and formats, the library community has joined the world of linked open data as a means of achieving better access and use of the data in its collections and of promoting reuse on the web. This article collects information about current library LOD services and projects, paying particular attention to authority catalogues by reviewing the bibliography published in recent years. Now, the Centre de Recursos per a l'Aprenentatge i la Investigació of the Universitat de Barcelona is joining the LOD community, offering its subject authority dataset, the Thesaurus of the University of Barcelona (THUB) in linked open data.

Keywords: Linked data, controlled vocabularies, open data, University of Barcelona Thesaurus.



Introducció

Aquest article és el resultat de l'anàlisi i el seguiment que, des de la Unitat de Procés Tècnic del Centre de Recursos per a l'Aprenentatge i la Investigació de la Universitat de Barcelona (CRAI-UB), s'ha dut a terme durant els darrers anys de les iniciatives de dades obertes enllaçades (LOD) en el nostre àmbit professional. El contingut se centra en els projectes al voltant dels catàlegs i, en especial, dels llenguatges controlats i de l'evolució de la normativa catalogràfica per a fer servir aquestes eines en el web semàntic. Al mateix temps ens ha servit per avalar la decisió d'implementar un primer projecte amb aquesta tecnologia. Es tracta de la publicació del Thesaurus de la Universitat de Barcelona (THUB) en LOD,¹ com a primera aportació de la nostra institució a aquest moviment.

Es tracta, doncs, d'una revisió de l'estat de la qüestió, tant des del punt de vista de la bibliografia com dels projectes en si, però sense intenció d'exhaustivitat, amb un interès especial en l'aplicació de les dades obertes enllaçades als vocabularis controlats i l'ús que se'n fa en la recuperació de la informació. Per a aprofundir en l'estat de la qüestió recomanem l'article de Mitchell, 2016. D'altra banda, volem deixar constància del fet que la nostra aproximació a aquest tema es fa a partir del punt de vista bibliotecari i no pas tecnològic. Amb aquestes premisses, hem intentat respondre a les preguntes que ens fèiem al voltant de què implica un projecte de LOD i a partir d'això hem seleccionat i analitzat els projectes que descrivim i que hem considerat de referència en aquest context.

El procés de reflexió que hem seguit és el que estructura aquest article. Després d'analitzar la situació actual i, en segon lloc, el context tecnològic i de canvi produït en les eines de re-

cuperació de la informació en els darrers anys, hem intentat identificar els principis en els quals ens hauríem de basar per a endegar un projecte en LOD. En tercer lloc, presentem una relació dels projectes que hem considerat més interessants, tant pel que fa a la creació de portals bibliogràfics com als projectes centrats en l'obertura de vocabularis controlats. Dels projectes destacats, allò que ens ha interessat ha estat analitzar els aspectes clau en la definició, l'abast i les fases de la implementació, els col·laboradors tecnològics que hi han participat, formats i tecnologies emprades, serveis oferts per a la proporció de dades i llicències d'ús aplicades a les dades, entre d'altres. Finalment, fem una breu presentació del nostre projecte destacant-ne els aspectes que tenen a veure amb els principis d'un projecte amb LOD.

Compartim amb tots els neòfits en el tema aquesta visió per a intentar entendre l'evolució que es preveu de la nostra tasca principal.

1. Context actual

Ningú posa en dubte que les biblioteques contenen i ofereixen en els seus catàlegs —bibliogràfics i d'autoritats, bases de dades, dipòsits, etc.— dades de qualitat generades intel·lectualment per personal especialitzat, tal com indica la Deutsche National Bibliothek² en els seus projectes de dades enllaçades. L'aportació del potencial de les dades de les biblioteques, arxius i museus al projecte del web semàntic es considera un component vertebrador per la qualitat i el tractament professional de les dades.³

Aquestes dades, fins fa poc, estaven tancades en les institucions i únicament estaven disponibles un cop s'entrava a les pàgines web. La reutilització d'aquestes dades ha estat fins ara

1. Thesaurus de la Universitat de Barcelona (THUB) <<http://crai.ub.edu/ca/recursos-d-informacio/thub>>.
2. Linked-Data-Service des Deutschen Nationalbibliothek <http://www.dnb.de/EN/Service/DigitaleDienste/LinkedData/linkedata_node.html>.
3. Joachim Neubert, Klaus Tochtermann, «Linked library data: offering a backbone for the semantic web» [en línia]. En: Knowledge Technology Week (3rd: 2011: Kajang, Malaysia). Lukose, Dickson; Ahmad, Abdul Rahim; Suliman, Azizah (eds.), *Knowledge Technology: KTW 2011: Communications in Computer and Information Science*, v. 295 (2012), Berlin; Heidelberg: Springer, 2012, p. 37-45, <https://doi.org/10.1007/978-3-642-32826-8_4> [Consulta: 08/11/2018].

El camí de la veritable evolució dels catàlegs i les eines de descoberta de la informació, en general, ha d'evolucionar cap a l'ús de les dades de qualitat enllaçades en obert, que es troben en els catàlegs bibliogràfics i d'autoritat.

feixuga i poc àgil i en molts casos ha obligat a duplicar les tasques de descripció i indexació encara que es treballi en el mateix format. D'aquí ve la importància de millorar el treball propi i en col·laboració mitjançant l'obertura d'aquestes dades de qualitat amb la finalitat de compartir-les i reutilitzar-les de manera fàcil, ràpida, econòmica, qualitativa i efectiva.⁴

Últimament es parla molt de les eines de descoberta que s'han creat amb l'objectiu de facilitar l'accés des d'una única finestra a tots els recursos que la biblioteca posa a disposició dels usuaris, tant les bases de dades com els dipòsits digitals i, en general, tots els recursos en línia oferts per les diferents empreses que en són creadores i distribuïdores. En alguns casos, les eines de descoberta tenen en compte els recursos oferts per les biblioteques a partir dels catàlegs bibliogràfics. Una vegada implementades aquestes eines, es posa en evidència que no aporten valor ni usen dades de qualitat, sinó que són mers recercadors al web a partir de paraules clau. Per tant, el camí de la veritable evolució dels catàlegs i les eines de descoberta de la informació, en general, ha d'evolucionar cap a l'ús de les dades

de qualitat enllaçades en obert, que es troben en els catàlegs bibliogràfics i d'autoritat,⁵ i que s'haurien de poder incorporar en les cerques d'aquestes eines de descoberta.

A partir de la literatura existent, es pot veure com ha evolucionat i com evoluciona encara, es materialitza i arrela en el món de les biblioteques el concepte i la realitat de les dades obertes enllaçades. Les diferents fases de desenvolupament dels projectes de LOD es poden resumir en una primera etapa de creació de la tecnologia que ha de permetre oferir les dades obertes, i una segona etapa en què s'inicien els projectes d'obertura de les dades per a enriquir-les i poder reutilitzar-les per a la creació d'altres projectes i productes en una tercera fase.

2. Dades obertes enllaçades i biblioteques

Tim Berners-Lee, l'inventor del World Wide Web, ja havia introduït el marc tecnològic de les dades enllaçades (*linked data*) l'any 2006, però no va ser fins al 2009,⁶ quan va demanar a tothom que compartís les seves dades, que el món no va començar a comprendre la idea que el web de dades començava a evolucionar cap al web semàntic. Des de l'aparició de les primeres iniciatives de publicació de metadades amb llicència oberta en tota mena de sectors governamentals i no-governamentals, comercials, socials, etc., hi ha hagut un creixement considerable de programes, projectes i producció acadèmica del tema.

4. Tiziana Possemato, «How RDA is essential in the reconciliation and conversion processes for quality linked data» [en línia], *JLIS.it: Italian Journal of Library, Archives, and Information Science*, v. 9, n. 1 (2018), <<https://www.jlis.it/article/view/12447>> [Consulta: 08/11/2018].
5. F. William Chickering; Sharon Q. Yang, «Evaluation and comparison of discovery tools: an update» [en línia], *Information Technology and Libraries*, v. 33, n. 2 (2014), p. 5-30, <https://ejournals.bc.edu/ojs/index.php/ital/article/view/3471/pdf_1> [Consulta: 08/11/2018]; Lorena Ávila García, Virginia Ortiz Repiso, David Rodríguez Mateos, «Herramientas de descubrimiento: ¿una ventanilla única?» [en línia]. *Revista española de documentación científica*, v. 38, n. 1 (2015): e077, <<https://doi.org/10.3989/redc.2015.1.1178>> [Consulta: 08/11/2018].
6. Tim Berners-Lee, «The next web of open, linked data» [en línia], *TED Talks*, feb. 2009, <https://www.ted.com/talks/tim_berners_lee_on_the_next_web.html> [Consulta: 08/11/2018].

El principi de les dades enllaçades es basa en l'oferiment de dades per a compartir-les, ampliar-les, utilitzar-les i reutilitzar-les en el context del web semàntic i en un format basat en identificadors, independents de l'idioma, que permeti funcionalitats multilingües. Aquestes característiques permeten descriure els recursos i enriquir els conjunts de dades de manera eficient i en col·laboració. L'escenari que es preveu amb la introducció de les LOD a les biblioteques possibilitarà la creació d'un nombre important de dades obertes per compartir que permetrà l'estalvi de les tasques repetitives que es fan en la catalogació. Al mateix temps, l'ús del web permetrà tenir les dades actualitzades i disponibles, cosa que facilitarà la identificació i citació de manera correcta i unívoca. En aquest context, les dades d'autoritat tenen un paper destacat perquè poden contribuir a reduir el nombre de descripcions redundants existents en el web i a construir un sistema de dades vinculades que doni suport a la normalització.

Aquests principis en què es basen les dades enllaçades són els mateixos en el web semàntic i queden recollits en l'informe final del W3C Library Linked Data Incubator Group.⁷ Les dades enllaçades s'han d'expressar segons una normativa que expliciti les relacions entre les coses, com ara el Resource Description Framework (RDF), i a partir del que coneixem com els Uniform Resource Identifiers (URI) (adreces web basades en identificadors estables i permanents). Aquest informe final extreu, després d'analitzar les iniciatives al voltant de les dades vinculades, un seguit de recomanacions adreçades principalment al nostre col·lectiu professional (directors de biblioteques, catalogadors, arxivers, etc.) des d'un punt de vista entenedor, que ha de servir com a referència per a introduir i entendre els projectes existents.

Les recomanacions per a dur a terme projectes de dades enllaçades que es desprenen de l'informe són:

- Identificar conjunts de dades que permetin assolir projectes més prioritaris i de menys cost, que puguin dur-se a terme sense pertorbar els serveis en marxa, com poden ser els fixers d'autoritats.
- Fomentar les dades obertes enfront dels drets de propietat en l'exposició de les dades enllaçades.
- Incrementar la participació de les biblioteques, especialistes en normalització, en la construcció del web semàntic a partir del desenvolupament de normes compatibles entre aquests dos àmbits i seguint codis de bones pràctiques.
- Preservar els conjunts d'elements i vocabularis de dades enllaçades i, per tant, els URI en què se sustenten, ja que són la clau del manteniment de l'ecosistema de dades vinculades.
- Aplicar l'experiència professional en el tractament i la preservació d'aquests conjunts de dades a llarg termini.

D'aquest mateix informe, en destaquem les consideracions següents:

- Hi ha un nombre petit d'empreses que es dediquen a desenvolupar sistemes integrats de biblioteques; per tant, les iniciatives estan molt condicionades per aquesta situació.
- Per aquest motiu, encara que n'hi ha un gran nombre, moltes iniciatives es troben en fase de prototip, ja que el dinamisme inicial té el seu contrapunt en l'estabilitat i disponibilitat a llarg termini.
- Per tant, en primer lloc cal identificar els projectes de LOD més prioritaris i de menys cost sense pertorbar els sistemes i serveis actuals per a poder ampliar i incrementar l'ús del web.

7. «Library Linked Data Incubator Group Final Report» [en línia], [Cambridge: The World Wide Web Consortium (W3C), date: 2011/10/25], <<https://www.w3.org/2005/incubator/ld/XGR-ld-20111025/>> [Consulta: 27/03/2018].

3. Desenvolupament de projectes de LOD en biblioteques

Quan es va començar a parlar del web semàntic, semblava un projecte inabastable i molt allunyat del que es feia en l'entorn de les biblioteques, els catàlegs bibliogràfics i les noves bases de dades emergents dels dipòsits digitals. En canvi, els termes *accés obert*, *dades de qualitat* i *treball en col·laboració* defineixen i defineixen conceptes totalment relacionats amb les tasques de les biblioteques i els bibliotecaris. No són, en cap cas, conceptes nous ni que ens resultin estranys. El que sí que resulta més allunyat de la nostra realitat és la manera d'oferir les dades de qualitat que gestionem en un entorn en què els grans cercadors i les xarxes socials prenen el protagonisme. En aquesta línia, es començà a parlar globalment de les dades enllaçades en obert de les biblioteques com a columna vertebral del web semàntic.⁸

Han passat deu anys d'ençà de la proposta de Tim Berners-Lee i l'escenari actual ja compta amb nombrosos projectes de biblioteques que intenten donar resposta a la seva petició. Molts d'aquests projectes encara estan en fase experimental i d'altres ja estan en funcionament oferint les dades enllaçades en obert i, per tant, en el marc del web semàntic.

Amb la voluntat de dur a terme aquests projectes de LOD, ha quedat patent que els formats de Machine-Readable Cataloging (MARC) majoritàriament usats en la gestió dels catàlegs bibliogràfics han quedat obsolets. Aquests han estat clau per a complir els objectius de les biblioteques de descriure, cercar i obtenir els documents en diferents suports i tipologies. Actualment, amb els canvis tecnològics que marquen una manera

El naixement de les tècniques de Recursos descripció i accés (RDA) marca una nova via de descriure i d'accedir als recursos d'informació en què predominen la identificació i les relacions de les dades per sobre de la descripció clàssica que es feia fins ara.

diferent d'obtenir la informació, aquests formats ja no assoleixen de manera eficaç els objectius marcats. Cal adaptar-se a nous formats i a altres sistemes de presentar i gestionar les dades que facilitin la participació en el web semàntic. Si fins ara la informació havia de ser compresa per les persones, ara, en aquest nou entorn, els nous formats han de permetre que les dades siguin compreses tant per les persones com per les màquines.⁹ El naixement de les tècniques de Recursos descripció i accés (RDA) marca una nova via de descriure i d'accedir als recursos d'informació en què predominen la identificació i les relacions de les dades per sobre de la descripció clàssica que es feia fins ara.¹⁰

3.1 Biblioteques nacionals

Les biblioteques nacionals de referència van començar a plantejar-se de manera seriosa projectes de LOD que es van materialitzar durant la dècada del 2010, com per exemple la Bibliothèque Nationale de France, la Biblioteca Nacional d'Espanya i la British Library, que tenen catàlegs col·lectius tan importants com l'Online Computer Library Center (OCLC) amb la col·laboració de la Library of Congress (LC), entre d'altres.¹¹

8. Joachim Neubert, Klaus Tochtermann, *op. cit.*

9. Andreu Sulé Duesa, «Schema.org: la millora de la visualització dels resultats en els cercadors i molt més» [en línia], *BiD: textos universitaris de biblioteconomia i documentació*, n. 34 (2015), <<http://dx.doi.org/10.1344/BiD2015.34.23>> [Consulta: 08/11/2018].

10. Tiziana Possemato, *op. cit.*

11. Ana Isabel Torre Bastida, Marta González Rodríguez, Esther Villar Rodríguez, «Datos abiertos enlazados (LOD) y su implantación en bibliotecas: iniciativas y tecnologías» [en línia], *El Profesional de la información*, v. 24, n. 2 (2015), p. 113-120, <<https://doi.org/10.3145/epi.2015.mar.04>> [Consulta: 08/11/2018].



Aquestes grans biblioteques han creat i estan desenvolupant en molts països serveis de dades enllaçades per a permetre, a llarg termini, que la comunitat web semàntica pugui emprar tot l'inventari de dades bibliogràfiques nacionals i de dades d'autoritat d'autors, títols i matèries que contenen. Els serveis de dades enllaçades treballen en el desenvolupament tècnic i l'optimització continuada de continguts i d'organització. Un dels objectius d'aquests serveis és garantir la fiabilitat i l'estabilitat del núvol en què es dipositaran aquestes dades. Les biblioteques nacionals, mitjançant aquests serveis, subministren les seves dades en l'estàndard Resource Description Framework (RDF), que permet als usuaris tornar a utilitzar-les d'una manera que no requereixi cap coneixement de formats bibliogràfics o d'autoritat.

Vegem-ne alguns exemples i les darreres novetats:

12. <<https://id.loc.gov/>>.

13. <<https://id.loc.gov/search/?q=&q=cs%3Ahttp%3A%2F%2Fid.loc.gov%2Fauthorities%2Fsubjects>>.

LIBRARY OF CONGRESS
LINKED DATA SERVICE

Library of Congress Linked Data Service, Authorities and Vocabularies

El servei de Linked Data Service de la Library of Congress¹² es va crear l'any 2009 després de la bona acollida que va tenir la publicació experimental de la *Library of Congress Subject Headings* (LCSH)¹³ l'any anterior. Publicada en llengua anglesa, la LCSH és, sens dubte, el vocabulari més emprat internacionalment per a l'establiment d'encapçalaments de matèries i, a la vegada, es considera el vocabulari comú per a establir vincles amb altres llistes d'encapçalaments en diferents idiomes. És per aquests dos motius que es va posicionar ràpidament com a font per a establir vincles amb altres vocabularis controlats d'indexació, la qual cosa va afavorir encara més el posicionament

d'aquesta institució dintre de la comunitat del web semàntic. Actualment el conjunt de dades de la LCSH conté més de set milions de termes RDF i està vinculat al conjunt de dades del Répertoire d'autorité-matière encyclopédique et alphabétique unifié (RAMEAU)¹⁴ de la Bibliothèque nationale de France, i al Gemeinsame Normdatei¹⁵ de la Deutsche Nationalbibliothek, entre d'altres.

El 2008, en un informe de la Library of Congress, s'indicava que el format MARC estava quedant desfasat per als estils de programació de l'època i l'any 2012 la Library of Congress anuncià el seu contracte amb Zepheira, una companyia de gestió de dades, per a desenvolupar una alternativa als formats MARC, basada en dades enllaçades. El mateix any la biblioteca anuncià un nou model anomenat MARC Resources (MARCR) i un esborrany molt més complet amb el nom de BIBFRAME.¹⁶

Actualment la Library of Congress Linked Data Service ofereix el seu conjunt de dades d'autoritat sense restriccions en l'ús, en gran varietat de formats, i continua publicant vocabularis addicionals i normes com ara els Preservation Datasets i les llistes i els codis del format MARC21.



Datos enlazados en la BNE

La col·laboració entre la BNE i el Grupo de Ingeniería Ontológica (Ontology Engineering Group, OEG) de la Universidad Politécnica de Madrid, iniciada el 2010, va fer possible la creació del portal de dades de la BNE.¹⁷ L'any

2011 es va obtenir un conjunt molt representatiu de dades enllaçades, que es considera la primera versió d'aquest portal.

Actualment hi ha disponible una nova versió del portal de dades que incorpora novetats i millores respecte a l'anterior, com per exemple la construcció d'una interfície visual amb un cercador per a facilitar l'explotació de les relacions entre les dades, la seva presentació i l'enriquiment amb dades de fonts externes; l'ampliació del rang i el nombre de registres bibliogràfics i d'autoritat, de manera que aquesta versió cobreix pràcticament tots els materials de la Biblioteca; la inclusió d'enllaços a exemplars digitalitzats a la Biblioteca Digital Hispànica, i la descripció de les dades mitjançant una ontologia pròpia de la BNE que reutilitza propietats de diferents vocabularis, per a desenvolupadors i experts.

La BNE va posar en marxa a mitjan 2017 el BNElab,¹⁸ com a espai virtual per al desenvolupament de productes i serveis a partir del contingut de la biblioteca, que va ser desenvolupat en col·laboració amb Red.es. Dins de BNElab es treballa per a millorar, estructurar, enriquir i posar a disposició pública el conjunt de dades de la BNE. Actualment la BNE permet la descàrrega i reutilització lliure de les dades del seu catàleg bibliogràfic i d'autoritats, en formats XML i MRC, del seu portal de dades enllaçades en format RDF mitjançant un punt d'accés SPARQL, i de la Biblioteca Digital Hispànica mitjançant el protocol OAI per a l'intercanvi i recollida, en formats MARCXML i Dublin Core. S'aplica la llicència d'ús Creative Commons Zero (CC0). Hi ha previst publicar nous conjunts de dades desagregables, més petits, que puguin analitzar-se amb eines digitals i sense coneixements d'SPARQL, i més formats (JSON, CSV). Actualment, en el seu repositori GitHub¹⁹ (plataforma de desenvolupament co-

14. <http://www.bnf.fr/fr/professionnels/anx_catalogage_indexation/a.referentiels_sujet.html>.

15. <http://www.dnb.de/EN/Standardisierung/GND/gnd_node.html>.

16. <<https://www.loc.gov/bibframe/pdf/marcld-report-11-21-2012.pdf>>.

17. <<http://datos.bne.es/inicio.html>>.

18. <<http://www.bne.es/bnelab/>>.

19. <<https://github.com/BNElab>>.



operatiu de programari) hi ha disponible el *Juego filarmónico*,²⁰ que és el primer projecte de reutilització dels fons digitals de la BNE.



British National Library



Rodoreda, Mercè, 1908-1983	
family name	Rodoreda
given name	Mercè
has created	Death in spring / Mercè Rodoreda Death in spring / Mercè Rodoreda Death in spring / Mercè Rodoreda In Diamond Square / Mercè Rodoreda In Diamond Square / Mercè Rodoreda In Diamond Square / Mercè Rodoreda In Diamond Square / Mercè Rodoreda The pigeon girl / Mercè Rodoreda War, so much war / Mercè Rodoreda
life event	1908 1983
name	Mercè Rodoreda
type	Agent Agent Person
is like	0000000121452645
same as	100258547

Figura 1. Exemple de cerca a la BNB.

La BNB Linked Data Platform²¹ es va crear l'any 2011, proporciona accés a la Bibliografia Nacional Britànica (BNB), publicada com a dades obertes vinculades, i està disponible a través de serveis SPARQL. El conjunt de dades és una font autoritzada d'informació sobre publicacions del Regne Unit des del 1950 fins a l'actualitat amb dades consistents i ben gestionades. Tots els registres tenen una classificació decimal de Dewey (DDC) i els encapçalaments estan sota control d'autoritats. Pel que fa a la vinculació a conjunts de dades externs, s'han

triat recursos generals per donar a les dades un context més ampli, com GeoNames, per país de publicació, i Lexvo per a idiomes. També enllaça amb els conjunts de dades del domini de la biblioteca: VIAF, LCSH, Dewey.info i l'etiqueta MARC corresponent al país.

Ofereix un subconjunt amb la Bibliografia Nacional Britànica completa. Inclou llibres publicats, incloent-hi monografies, publicacions en sèrie i llibres actuals, que representen aproximadament 3,9 milions de registres.



Linked-Data-Service der Deutschen Nationalbibliothek

El Servei de dades enllaçades de la Biblioteca Nacional Alemanya²² proveeix les seves dades en format d'exportació RDF des de l'any 2010. La llicència d'ús Creative Commons Zero (CC0 1.0) s'aplica al fitxer d'autoritat integrat Gemeinsame Normdatei (GMD), que és l'índex estandaritzat de control d'autoritats de noms de persona, institucions, congressos, títols, noms geogràfics i encapçalaments de matèria que la Biblioteca Nacional d'Alemanya ha emprat internacionalment en els països de parla alemanya (Alemanya, Suïssa i Àustria), i als registres bibliogràfics del Servei de dades enllaçades.

D'ençà de finals de gener de 2014, algunes dades d'autoritat geogràfica contenen coordenades sense modificacions de la base de dades GeoNames²³ i des de l'octubre de 2015, algunes dades bibliogràfiques contenen classes temàtiques de Thema,²⁴ un esquema de categoria de termes per a un comerç global de llibres.

20. <<http://juegofilarmonico.bne.es/>>.

21. <<http://bnb.data.bl.uk/>>.

22. <http://www.dnb.de/EN/Service/DigitaleDienste/LinkedData/linkedd_data_node.html> (en anglès).

23. <<http://www.geonames.org/>>.

24. <<https://www.editeur.org/151/Thema/>>.

Per a convertir les dades a RDF s'utilitza Metafaçade, desenvolupada per culturegraph.org. Les dades de prova en representació RDF (RDF / XML o Turtle) estan disponibles, de manera gratuïta, en visualitzar un registre complet en el Katalog der Deutschen Nationalbibliothek,²⁵ dins el menú «Aktionen» (que apareix a la part dreta de la pantalla) o a la Datenschop.²⁶ Tots els formats s'ofereixen en UTF-8.

{BnF

Bibliothèque nationale de France

Una característica central del projecte data.bnf.fr²⁷ és la seva conformitat amb els estàndards web semàntics. Ofereix dos avantatges de les LOD: la possibilitat d'identificar inequívocament els registres bibliogràfics i d'autoritats, i, en segon lloc, fa possible l'expressió estandarditzada de les relacions entre conjunts de dades. Es creen relacions d'equivalència entre els identificadors de la Bibliothèque nationale de France (BnF) i entitats equivalents de la Biblioteca Nacional d'Espanya (BNE), o entre un lloc identificat pel BnF i el mateix lloc per a l'Institut National de l'information Géographique et Forestière (IGN) francès.

El lloc web permet agrupar en una mateixa pàgina de resultats sobre l'autor, les obres, les matèries, les dades i els llocs, dels recursos de la Biblioteca Nacional de França, així com els recursos externs. En aquestes pàgines s'articulen els diferents continguts, enllaços i serveis que la BnF proporciona a internet, que per raons tècniques es dispersen en les diferents aplicacions que té. S'inicià el juliol de 2011 i està en evolució constant.

25. <http://www.dnb.de/EN/Home/home_node.html>.

26. <<https://portal.dnb.de/metadataShop.htm>>.

27. <<http://data.bnf.fr/>>.

28. <<http://www.share-vde.org/>>.

3.2. Projectes individuals i/o conjunts de biblioteques universitàries, especialitzades i nacionals, arxius i museus

A part de les biblioteques nacionals, les biblioteques universitàries també desenvolupen projectes amb entorns de descoberta virtual alguns dels quals es basen en els nous models dels requisits funcionals dels registres bibliogràfics (FRBR) i el format BIBFRAME.

Són projectes d'una única institució o organització, o d'un grup d'institucions amb vincles geogràfics o altres característiques. Alguns d'aquests projectes afecten la totalitat de les dades dels catàlegs i d'altres només n'afecten una part. Entre aquests també hem citat projectes que tenen com a punt de partida els vocabularis controlats de matèries o d'autoritats. També cal destacar que alguns d'aquests projectes estan en una fase inicial o beta i d'altres ja estan plenament consolidats.



The Share-VDE PROJECT

En l'àmbit catalogràfic cal destacar The Share-VDE Project,²⁸ en què participen institucions de l'àmbit anglosaxó tan importants com: Stanford University, University of California Berkeley, Yale University, Library of Congress, National Library of Medicine, University of Chicago, University of Michigan Ann Arbor, Harvard University i Massachusetts Institute of Technology, entre d'altres. Segons Possemato (2018), aquest projecte il·lustra perfectament la fita que es vol assolir una vegada finalitzat el procés d'evolució catalogràfica dut a terme en els últims anys, amb la publicació en RDF

de les dades catalogràfiques segons el model FRBR i l'aplicació del format BIBFRAME, que substituïria el format tancat MARC.

En aquest moment, s'ha iniciat una primera fase que inclou les dades enregistrades en 2.308.204 registres bibliogràfics i 3.601.327 registres d'autoritats.

El projecte preveu un seguit d'activitats addicionals i el processament per a l'enriquiment i la conversió d'uns cent milions de registres de catàleg durant la fase 2 del projecte.



Europeana

En l'àmbit bibliogràfic, documental i museístic també es duen a terme moltes iniciatives. Cada cop hi ha més biblioteques, arxius i museus que treballen per aportar la seva contribució a la infraestructura d'informació global del web semàntic. Cal destacar el portal europeu del patrimoni cultural Europeana²⁹ com a projecte comunitari de dades obertes enllaçades. Europeana, a part d'enllaçar diferents representacions de conceptes relacionats en el web semàntic, permet a les biblioteques treballar amb les seves dades, i enriquir les pròpies metadades a la vegada que també s'arriba a molts més usuaris. D'aquesta manera, compleix un dels seus objectius principals, el foment del patrimoni cultural per al públic en general i, d'altra banda, també dona visibilitat a les biblioteques, arxius i museus que participen en aquest portal digital europeu.

Europeana conté i dona accés a un total de 58 milions d'obres d'art, objectes, llibres, vídeos i sons d'arreu d'Europa.



LODLAM (Linked Open Data in Libraries Archives and Museums)

La comunitat de persones, tècnics i professionals interessats a treballar en projectes de dades enllaçades en biblioteques, arxius i museus ha creat LODLAM,³⁰ una xarxa informal d'intercanvi de recursos i de col·laboració amb altres persones interessades o que treballen amb dades obertes vinculades en el seu àmbit.

LODLAM està dissenyat per ser un lloc central per a connectar-se i compartir notícies, preguntes, recursos, presentacions i articles relacionats amb els projectes de LOD que es duen a terme amb especial èmfasi en biblioteques, arxius i museus.



Portale Linked Open Data del CoBiS

El projecte pilot Linked Open Data del Coordinamento Biblioteche Speciali e Specialistiche di Torino (CoBiS³¹), realitzat gràcies a un acord amb la Regió del Piemont, té com a objectiu la creació de vincles entre les dades bibliogràfiques i d'autoritats de les biblioteques participants en l'experiència i la comunitat web.

Des d'un únic portal es pot accedir a les dades procedents dels catàlegs de les institucions participants encara que s'utilitzin diferents programes de gestió. Aquesta confluència permet enriquir les dades resultants i també possibilita la creació d'enllaços externs.

És un projecte fet amb la mateixa filosofia, imatge i configuració que el portal Datos.bne.es de la BNE.

29. <<https://www.europeana.eu/portal/ca>>.

30. <<http://lod-lam.net/>>.

31. <<http://dati.cobis.to.it>>.



FUNDACIÓN BIBLIOTECA VIRTUAL
MIGUEL DE CERVANTES

Biblioteca Virtual Miguel de Cervantes

El catàleg de la Biblioteca Virtual Miguel de Cervantes³² conté uns 200.000 registres bibliogràfics i d'autoritats creats originàriament d'acord amb la norma MARC21. Les entrades del catàleg han estat migrades cap a una nova base de dades relacional. El model de dades s'adhereix als models conceptuals promoguts per l'IFLA, en particular, a les especificacions FRBR i FRAD. El contingut de la base de dades ha estat posteriorment assignat, mitjançant un procediment automatitzat, a ternes RDF que utilitzen bàsicament el vocabulari RDA per descriure les entitats, així com les seves propietats i relacions. Gràcies a la seva naturalesa de dades obertes, aquestes són públiques i es poden unir i utilitzar fàcilment per a noves aplicacions creades per desenvolupadors i institucions externes.³³

Proporciona enllaços al catàleg tradicional, com ara cerca de contingut o portals temàtics. Inclou enllaços a bases de dades externes com ara VIAF, DBpedia i l'International Standard Name Identifier (ISNI) i també altres tipus d'enllaços com ara formularis de cerca amb els camps prefixats: per exemple, Europea o Wikipedia. També inclou enllaços a la Viquipèdia per autors, idiomes i dates, i enllaços a la Library of Congress per a la codificació dels idiomes.



DataUPSA de la Universidad Pontificia de Salamanca

OpenBiblio, el Portal de dades obertes de la biblioteca de la Universitat Pontificia de Salamanca³⁴ és un projecte de publicació de dades de personalitats i obres de la Universitat Pontificia de Salamanca amb dades obertes enllaçades.

El projecte té com a principal objectiu publicar informació bibliogràfica que proporcioni dades fidedignes de personalitats locals a partir dels registres d'autoritat, així com dades bibliogràfiques del catàleg de la biblioteca. El model de disseny és aplicable a qualsevol entorn de biblioteca universitària i/o especialitzada, tant des del punt de vista del sistema d'informació o programari que es presenta, com de les seves ontologies.³⁵



Linked Jazz project

Linked Jazz³⁶ és un projecte de recerca que investiga l'aplicació de les tecnologies de dades obertes enllaçades als materials digitals del patrimoni cultural. Té com a objectius:

- Descobrir connexions significatives entre documents i dades relacionades amb la vida personal i professional dels artistes de jazz.
- Desenvolupar eines i mètodes àmpliament aplicables per treballar amb dades obertes enllaçades.

32. <<http://data.cervantesvirtual.com/>>.

33. Gustavo Candela, *et al.*, «Migration of a library catalogue into RDA linked open data» [en línia], *Semantic Web*, v. 9, n. 4 (2018), p. 481-491, <<https://doi.org/10.3233/SW-170274>> [Consulta: 08/11/2018].

34. <<http://dataupsa.upsa.es/>>

35. Ana María Feroso García, *et al.*, «Sistema de modelado semántico para catalogación, clasificación, consulta y publicación en abierto de información bibliográfica» [en línia], *El Profesional de la información*, v. 27, n. 2 (2018), p. 410-418, <<https://doi.org/10.3145/epi.2018.mar.20>> [Consulta: 08/11/2018].

36. <<https://linkedjazz.org/>>.



El projecte es basa en materials d'història del jazz en format digital per a exposar les relacions entre músics i revelar la seva xarxa comunitària. Els nous modes de connexió de dades culturals tenen el potencial d'obrir noves vies d'investigació i d'implicació de la comunitat sense precedents.

Per crear les LOD, s'han desenvolupat un conjunt d'eines: un analitzador de transcripcions, una eina que assigna noms, una eina de conservació i una eina de proveïment participatiu (*crowdsourcing*). Aquestes eines funcionen conjuntament per trobar els noms esmentats durant la interrogació per assignar una identificació positiva a cadascun, també desambigüen noms utilitzant recursos en línia com DBpedia i VIAF. L'analitzador de transcripcions també reconeix l'estructura de preguntes i respostes de la història oral. Com que les persones són esmentades com a resultat d'una interrogació, les ternes RDF simples entre els interrogats i les persones esmentades es creen sota la forma de «know of».

VIAF

VIAF: The Virtual International Authority File

El servei VIAF³⁷ proporciona a les biblioteques i als usuaris de la biblioteca l'accés als principals catàlegs d'autoritat de noms. Els col·laboradors de VIAF subministren regularment dades d'autoritats que coincideixen amb VIAF, enllaços i grups. Totes les descripcions d'una entitat determinada es fusionen en un clúster que agrupa els diferents noms d'aquesta entitat. Aquest servei permet als usuaris identificar noms, ubicacions, obres i expressions, tot conservant les preferències regionals per a l'idioma, l'ortografia i l'escriptura.

Així, de manera similar a les monedes que es ponderarien contra els seus equivalents en un

país estranger, els identificadors i enllaços entre conjunts de dades, permeten comparar els conjunts de metadades relacionats amb objectes comuns: aquest tipus de comparació revela la riquesa o la pobresa de formats de dades, precisió o inexactitud de descripcions o, fins i tot, diferències en les polítiques adoptades de catalogació. Un exemple d'això es pot trobar al servei de l'Autoritat Virtual (VIAF), que funciona com un centre d'identificació i promou la vinculació de registres d'autoritats idèntics a tot el món.³⁸



Biblioteca Virtual de Polígrafos de la Fundación Ignacio Larramendi

Les biblioteques virtuals de la Fundación Ignacio Larramendi s'agrupen amb la denominació Biblioteca Virtual de Polígrafos,³⁹ que reuneix uns 900 polígrafs espanyols, portuguesos, hispanoamericans i brasilers (entre els quals hi ha escriptors de ciència, exploracions, literatura, pensament, economia i altres ciències) i una sèrie de biblioteques virtuals temàtiques i d'autor que aniran augmentant i en les quals s'agrupen els polígrafs i les obres relacionades.

Es pot consultar com un conjunt, però també mitjançant les biblioteques virtuals temàtiques o d'autor que la componen. És a dir, l'evolució del projecte ha donat lloc al fet que el marc general de la Biblioteca Virtual de Polígrafos s'hagi anat detallant en diferents biblioteques virtuals, que unifiquen en el temps i en l'espai l'activitat dels pensadors espanyols, hispanoamericans, portuguesos i brasilers.

En l'actualitat la Biblioteca Virtual de Polígrafos és un exemple de com un projecte bibliogràfic pot encaminar —i ho ha de fer— la reutilització de dades i de continguts oberts de tota mena.

37. <<https://viaf.org/>>.

38. Raphaëlle Lapôtre, «Library metadata on the web: the example of data.bnf.fr» [en línia], *JLIS.it: Italian Journal of Library, Archives, and Information Science*, v. 8, n. 3 (2017), p. 58-70, <<https://www.jlis.it/article/view/12402>> [Consulta: 08/11/2018].

39. <<http://www.larramendi.es/i18n/inicio/inicio.do>>.

Aquest projecte, gràcies al seu plantejament, es troba, pel seu nom en anglès (Polymath Virtual Library), entre els casos d'ús del W3C Library Linked Data Incubator Group i d'Europeana Data Model.



FAST (Faceted Application of Subject Terminology)

FAST és un projecte de l'OCLC Research en col·laboració amb la Library of Congress que pretén desenvolupar encapçalaments de matèria simplificats per a les metadades i crear un nou tesaurus basat en els Library of Congress Subject Headings (LCSH) per tal de facilitar la descripció del contingut dels recursos electrònics. FAST va començar a crear i desenvolupar a finals de 1998 i els seus principals objectius són la compatibilitat amb metadades existents, la facilitat d'assignació de matèries, l'efectivitat de recuperació, el cost de manteniment i la interoperabilitat semàntica.⁴⁰

Els termes del vocabulari FAST es divideixen en vuit categories o facetes: noms personals, noms corporatius, noms geogràfics, esdeveniments, títols, períodes de temps i termes de gènere/forma. FAST adopta un format modular en què cada faceta disposa de grups discrets d'epígrafs precoordïnats en què s'empren encapçalaments amb subdivisions que pertanyen a la mateixa faceta; i també postcoordinats perquè, en els casos en què no s'han desenvolupat encapçalaments complexos o en els quals els encapçalaments pertanyen a

diferents facetes, no s'empren subdivisions, sinó que s'estableixen encapçalaments addicionals.⁴¹ Actualment el fitxer d'autoritat FAST conté més d'1.700.000 registres d'autoritat de matèria.

FAST Linked Data⁴² està disponible amb llicència Open Data Commons. També es pot descarregar amb aquesta llicència a la Faceted Application of Subject Terminology Dataset (FAST).⁴³

En el Fòrum electrònic de l'Association for Library Collections & Technical Services, de l'American Library Association (ALCTS e-Forum)⁴⁴ es pot trobar més informació dels possibles usos de FAST per a catalogar i descobrir recursos i implementar-los a les biblioteques.



The Getty Vocabularies

*The Getty Vocabularies*⁴⁵ són vocabularis controlats que contenen terminologia multilingüe per a descriure recursos relacionats amb l'art visual, l'arquitectura, l'arqueologia, la conservació i el patrimoni cultural.⁴⁶ Actualment estan disponibles en dades obertes enllaçades per poder compartir-les en el web semàntic i, a més, tots els recursos estan vinculats entre si. Els vocabularis inclouen: l'Art & Architecture Thesaurus (AAT), el Getty Thesaurus of Geographic Names (TGN), publicats tots dos en dades obertes enllaçades l'any 2014; l'Union List of Artist Names (ULAN), publicat en LOD

40. Rebecca J. Dean, «FAST: development of simplified headings for metadata» [en línia], *Cataloging & Classification Quarterly*, v. 39, n. 1-2 (2004), p. 331-352, <https://doi.org/10.1300/J104v39n01_03> [Consulta: 08/11/2018].

41. Marilyn Montalvo Montalvo, «LCSH, FAST y Delicious: vocabularios normalizados y nuevas formas de catalogación temática» [en línia], *Anales de documentación*, v. 14, n. 1 (2011), <<http://revistas.um.es/analesdoc/article/view/120141>> [Consulta: 08/11/2018].

42. <<http://id.worldcat.org/fast/>>.

43. <<http://www.oclc.org/research/activities/fast/download.htm>>.

44. <<http://www.ala.org/alctsnews/features/fast-cataloging>>.

45. <<http://www.getty.edu/research/tools/vocabularies/lod/>>.

46. Patricia Harpring, «Linked open data in the cultural heritage world: issues for information creators and users» [en línia]. En: *CLIR connect*, [Washington: Council on Library and Information Resources, posted 03/20/2014], <<https://connect.clir.org/blogs/patricia-harpring/2014/03/20/linked-open-data-in-the-cultural-heritage-world-issues-for-information-creators-and-users>> [Consulta: 08/11/2018].

el 2015; el Cultural Objects Name Authority (CONA) i la Iconography Authority (IA), que és un mòdul dintre de CONA. Actualment s'està treballant amb el mapatge de CONA a CIDOC CRM i en el desenvolupament de la *Iconography Authority*. L'AAT, TGN i ULAN estan disponibles en LOD, es publiquen en els formats SKOS i SKOS-XL i amb llicència Open Data Commons Attribution License (ODC-By) 1.0, que permet compartir, crear i adaptar les dades amb atribució.



Lista de Encabezamientos de Materia para las Bibliotecas Públicas (LEMBP)

La LEMBP⁴⁷ es va publicar com a dades obertes enllaçades el mes de juliol de 2011. Els encapçalaments de matèria es troben vinculats a LCSH i, actualment, s'està treballant per vincular-los a RAMEAU i Schlagwortnormdatei (SWD). La vinculació a aquestes tres llistes permetrà actualitzar-los, ja que des de la primera edició, el 1986, només s'havia actualitzat en la segona edició revisada de 1993.

Dins el mateix projecte també es van traduir al format SKOS la Llista d'encapçalaments de matèria de la Biblioteca de Catalunya (LEMAC), la Lista de encabezamientos de materia en galego (LEMAG) i la llista bilingüe Encabezamientos de materia en vasco (LEMAV).

Els encapçalaments d'aquest conjunt de fitxers d'autoritats poden ser recuperats mitjançant un cercador simple, un d'avançat o un *endpoint* SPARQL. També es poden descarregar els fitxers complets tant en format MARC21 com en SKOS (RDF/XML) per a utilitzar-los i reutilitzar-los. La LEMBP inclou 19.152 con-

ceptes en castellà; la LEMAC, 37.480 en català; la LEMAG, 4.954 en gallec, i la LEMAV, 114.172 conceptes en castellà i eusquera.

4. Thesaurus de la UB, un pas dins de les LOD



Després d'aquest procés de reflexió, inspirats per les iniciatives exposades i altres experiències, vam valorar que era oportú dur a terme el nostre propi projecte de dades enllaçades. Ho vam considerar per diverses raons:

- Disposàvem d'un conjunt de dades que permetia una primera aproximació al tema sense alterar el funcionament dels serveis actuals, amb caràcter d'originalitat i que volíem compartir. Es tracta del Thesaurus de la UB, elaborat per nosaltres des de fa més de 25 anys, en català, multidisciplinari i amb previsions de ser multilingüe.
- Volíem millorar-ne la visualització de cara als usuaris i, per tant, buscàvem una nova eina tecnològica que ens permetés obrir les possibilitats del tesaurus com a eina de recuperació.

La descripció més detallada del projecte es va presentar a la secció d'experiències de les 15es Jornades Catalanes d'Informació i Documentació,⁴⁸ juntament amb l'empresa amb què hem desenvolupat el projecte, Nubilum, quan encara estava en una fase d'implementació; per tant, ens vam centrar més en la descripció del projecte tecnològic. En l'actualitat, el projecte és visible i ha satisfet les nostres expectatives inicials. Per això, volem tancar aquest

47. <<http://id.sgcb.mcu.es/>>.

48. Rosa Fabeiro, Pep Casals. «El Thesaurus de la Universitat de Barcelona: 25 anys en constant evolució» [en línia]. En: *Jornades Catalanes d'Informació i Documentació* (15es: 2018: Barcelona), *15es Jornades Catalanes d'Informació i Documentació, 10 i 11 de maig de 2018*, [Barcelona: COBDC, 2018], <<http://www.cobdc.net/15JCID/wp-content/uploads/2018/05/Exp17.pdf>> [Consulta: 08/11/2018].

procés de reflexió destacant, malgrat les diferències, els aspectes que el nostre projecte té en comú amb les iniciatives recollides en aquest article i que estan en consonància amb els principis i les recomanacions internacionals en aquest àmbit. L'aplicació d'aquests principis ha suposat reptes i decisions que també volem compartir:

- Preservació dels Uniform Resource Identifier (URI). Un dels primers obstacles va ser garantir la preservació dels URI dels termes que configuren el nostre vocabulari. No només com una qüestió tècnica sinó també de procediment. El nostre tesaurus, fet a mida per a la indexació dels documents de les nostres col·leccions, no preveia fins ara mantenir el registre històric de canvis produïts. Per tant, s'optava per eliminar termes quan quedaven orfes de documents. Actualment es mantenen els termes independentment de l'ús en el nostre sistema d'informació per a preservar els conjunts d'elements i, per tant, els URI en què se sustenten.
- Foment de les dades obertes amb l'aplicació de la llicència Creative Commons Zero (CC0) per al nostre vocabulari, majoritàriament utilitzada en els casos estudiats, que està en línia amb la política de la nostra institució i cerca afavorir l'ús i l'intercanvi.
- Tria de l'ontologia SKOS, basada en el format RDF, com a model per a la representació del nostre vocabulari. Es tracta d'un model pensat per a la representació de l'estructura i el contingut d'esquemes conceptuals i de qualsevol tipus de vocabulari controlat. Tot i que el vocabulari es gestiona des del nostre SIGB en format MARC, es fa un mapatge d'aquest model de representació RDF que permet obrir les dades per a l'intercanvi i la reutilització, i això ens permet crear un entorn més amigable i entenedor per als usuaris.
- Desplegament del tesaurus en format multilingüe amb la traducció d'etiquetes equivalents. La tecnologia utilitzada ens ha permès generar les versions multilingües de

La nova tecnologia també ens ha aportat millores en la gestió de la coherència en les relacions establertes entre els termes i la detecció de manca de baules en alguns dels arbres temàtics. El canvi filosòfic ens ha fet treballar en les relacions i les jerarquies amb una visió més àmplia, per oferir una eina terminològica més oberta a la comunitat.

manera més fàcil i consistent. Actualment es pot consultar en català, castellà, anglès i francès.

- Incorporació d'una visualització gràfica que permet la navegació entre els diferents nodes i explicita les relacions i vinculacions entre els termes d'una manera més dinàmica i intuïtiva.
- Accés lliure a les dades obertes del THUB. El CRAI-UB permet la descàrrega i utilització del nostre vocabulari controlat en format RDF, a partir d'un *endpoint* de consulta SPARQL i un punt de descàrrega dels fitxers en format RDF/XML i N-Triples.

Thesaurus de la Universitat de Barcelona (THUB)

TESAURUS	Matèries
TERME PREFERENT	Dades enllaçades
TERMES NO PREFERENTS	Dades enllaçables, Dades obertes, Dades obertes enllaçables, Dades obertes enllaçades, LOD (Linked open data), Linked open data
TERMES GENÈRICS	Metadades, Web semàntic
MICROTESAURUS	Informàtica, Biblioteconomia i documentació
TERMES EQUIVALENTS	Datos enlazados [BNE/A], Linked data [LCSH]
ENLLAÇ PERMANENT	http://crai.ub.edu/thub/concept/a13492500
CERCA DOCUMENTS	CERCA DB, RPÒSIT UB, RECERCA

Figura 2. Doble visualització dels termes al THUB.



La nova tecnologia també ens ha aportat millores en la gestió de la coherència en les relacions establertes entre els termes i la detecció de manca de baules en alguns dels arbres temàtics. El canvi filosòfic ens ha fet treballar en les relacions i les jerarquies amb una visió més àmplia, deslligada de les dependències dels termes amb els documents de la nostra col·lecció, per oferir una eina terminològica més oberta a la comunitat.

Al mateix temps, hem detectat noves vies d'actuació per a continuar creixent i desenvolupant altres funcionalitats que ens semblen útils per als nostres usuaris. Per exemple, incrementar les relacions amb altres tipologies de documents de les nostres col·leccions (fotografies, documents d'arxiu, gràfics...) descrits en diferents contextos, a partir de la indexació, o interconnectar els diferents vocabularis controlats que utilitzem; per tant, obrir el nostre catàleg Autoritats UB⁴⁹ en RDF, entre d'altres.

Som conscients que encara hi ha molta feina per fer en aquesta línia d'obertura i enllaç de les dades de qualitat que elaborem en el món bibliotecari, però aquesta primera aproximació entorn de les dades enllaçades ens ha obert noves possibilitats de millora i de participació en la construcció del web semàntic.

5. Conclusions

Les biblioteques són serveis creats per satisfer les necessitats d'informació actuals i futures d'una comunitat d'usuaris. Per poder dur a terme aquesta missió han d'emprar les millors eines que la tecnologia els pugui oferir en cada moment i adaptar-se a les noves necessitats.

Un dels fonaments ideològics de la cultura de les LOD és la creença que servir la comunitat significa obrir la més gran quantitat de serveis, projectes, i dades útils i de qualitat, fent-los no només accessibles a través del World Wide Web, sinó essent part integrant del World Wide Web.

A través d'internet, les xarxes socials, els portals i els cercadors, la informació arriba cada cop a més usuaris. Per això, obrir les dades, vincular-les i intercanviar-les és una missió que afavoreix les biblioteques en molts sentits: en mostra els continguts i fons a la xarxa i les ajuda a controlar i augmentar la rellevància que tenen creant nous fluxos de treball, aprofitant la utilització i la reutilització de les dades i observant les tendències d'ús dels seus recursos, entre d'altres.

Familiaritzar els professionals i els responsables de les biblioteques amb els avantatges i el potencial de la tecnologia de les LOD no ha estat fàcil i ha requerit inversions inicials costoses, però amb el temps s'està demostrant que, en obrir les dades, les biblioteques estan creant beneficis per si mateixes, no només pel que fa a reducció de costos econòmics i de repetició de tasques sinó també, i més important, amb la generació de beneficis per als usuaris d'internet d'avui i de demà. S'ha fet molta feina i hi ha hagut molts canvis: noves tecnologies, nous models, noves normes i avenços cap als nous formats. Tot això per facilitar que les dades siguin compreses tant per les persones com per les màquines. Tot un esforç intel·lectual dels professionals de les biblioteques per aconseguir dades consistents, de qualitat, accessibles i reutilitzables. Les eines de descoberta no ens ajuden a portar a terme de manera satisfactòria la nostra missió, per tant, la tecnologia de LOD ens mostra el camí per satisfer les demandes dels nostres usuaris a partir de les eines que ells mateixos coneixen i utilitzen.

Encara queda feina per fer en el camp de l'alfabetització de bibliotecaris i usuaris en l'ús i la reutilització de les dades, i també es necessita més suport informàtic per poder avançar més ràpidament en l'obertura de dades. Malgrat això, en pocs anys s'han desenvolupat milers de projectes de LOD de tot tipus i el núvol de dades obertes enllaçades rellevants per a biblioteques no para de créixer.

49. <<http://www.bib.ub.edu/fileadmin/autoritats/>>.

Bibliografia

ÁVILA GARCÍA, Lorena; ORTIZ REPISO, Virginia; RODRÍGUEZ MATEOS, David. «Herramientas de descubrimiento: ¿una ventanilla única?» [en línia]. *Revista española de documentación científica*, v. 38, n. 1 (2015): e077. <<https://doi.org/10.3989/redc.2015.1.1178>> [Consulta: 08/11/2018].

BERMES, Emmanuelle; BOULET, Vincent; LECLAIRE, Céline. «Améliorer l'accès aux données des bibliothèques sur le web: l'exemple de data.bnf.fr» [en línia]. En: IFLA World Library and Information Congress (2016: Columbus, OH). *Connections. Collaboration. Community*. Session 81 - Division III Programme. <<http://library.ifla.org/1447/1/081-bermes-fr.pdf>> [Consulta: 08/11/2018].

BERNERS-LEE, Tim. «The next web of open, linked data» [en línia]. *TED Talks*, feb. 2009. <https://www.ted.com/talks/tim_berners_lee_on_the_next_web.html> [Consulta: 08/11/2018].

CANDELA, Gustavo, *et al.* «Migration of a library catalogue into RDA linked open data» [en línia]. *Semantic Web*, v. 9, n. 4 (2018) p. 481-491. <<https://doi.org/10.3233/SW-170274>> [Consulta: 08/11/2018].

CHICKERING, F. William; YANG, Sharon Q. «Evaluation and comparison of discovery tools: an update» [en línia]. *Information Technology and Libraries*, v. 33, n. 2 (2014), p. 5-30. <https://ejournals.bc.edu/ojs/index.php/ital/article/view/3471/pdf_1> [Consulta: 08/11/2018].

DEAN, Rebecca J. «FAST: development of simplified headings for metadata» [en línia]. *Cataloging & Classification Quarterly*, v. 39, n. 1-2 (2004), p. 331-352. <https://doi.org/10.1300/J104v39n01_03> [Consulta: 08/11/2018].

DELIOT, Corine. «Publishing the British National Bibliography as linked open data» [en línia]. *Catalogue & index*, n. 174 (2014), p. 13-18. <http://www.bl.uk/bibliographic/pdfs/publishing_bnb_as_lod.pdf> [Consulta: 08/11/2018].

FABEIRO, Rosa; CASALS, Pep. «El Thesaurus de la Universitat de Barcelona: 25 anys en constant evolució» [en línia]. En: Jornades Catalanes d'Informació i Documentació (15es: 2018: Barcelona). *15es Jornades Catalanes d'Informació i Documentació, 10 i 11 de maig de 2018*. [Barcelona: COBDC, 2018]. <<http://www.cobdc.net/15JCID/wp-content/uploads/2018/05/Exp17.pdf>> [Consulta: 08/11/2018].

FERMOSO GARCÍA, Ana María, *et al.* «Sistema de modelado semántico para catalogación, clasificación, consulta y publicación en abierto de información bibliográfica» [en línia]. *El Profesional de la información*, v. 27, n. 2 (2018), p. 410-418. <<https://doi.org/10.3145/epi.2018.mar.20>> [Consulta: 08/11/2018].

Harpring, Patricia. «Linked open data in the cultural heritage world: issues for information creators and users» [en línia] En: *CLIR connect*. [Washington: Council on Library and Information Resources, posted 03/20/2014]. <<https://connect.clir.org/blogs/patricia-harpring/2014/03/20/linked-open-data-in-the-cultural-heritage-world-issues-for-information-creators-and-users>> [Consulta: 08/11/2018].

LAPÔTRE, Raphèle. «Library metadata on the web: the example of data.bnf.fr» [en línia]. *JLIS.it: Italian Journal of Library, Archives, and Information Science*, v. 8, n. 3 (2017), p. 58-70. <<https://www.jlis.it/article/view/12402>> [Consulta: 08/11/2018].

«LIBRARY Linked Data Incubator Group Final Report» [en línia]. [Cambridge: The World Wide Web Consortium (W3C), date: 2011/10/25]. <<https://www.w3.org/2005/incubator/ld/XGR-ld-20111025/>> [Consulta: 08/11/2018].

MITCHELL, Erik T. «Library linked data: early activity and development» [en línia]. *Library Technology Reports*, v. 52, n. 1 (2016). <<https://journals.ala.org/ltr/issue/view/534>> [Consulta: 08/11/2018].

MONTALVO MONTALVO, Marilyn. «LCSH, FAST y Deliculous: vocabularios normalizados y nuevas formas de catalogación temática» [en línia]. *Anales de documentación*, v. 14, n. 1 (2011). <<http://revistas.um.es/analesdoc/article/view/120141>> [Consulta: 08/11/2018].

NEUBERT, Joachim; TOCHTERMANN, Klaus. «Linked library data: offering a backbone for the semantic web» [en línia]. En: Knowledge Technology Week (3rd: 2011: Kajang, Malaysia). Lukose, Dickson; Ahmad, Abdul Rahim; Suliman, Azizah (eds.). *Knowledge Technology: KTW 2011: Communications in Computer and Information Science*, v. 295 (2012). Berlin; Heidelberg: Springer, 2012, p. 37-45. <https://doi.org/10.1007/978-3-642-32826-8_4> [Consulta: 08/11/2018].

POSSEMATO, Tiziana. «How RDA is essential in the reconciliation and conversion processes for quality linked data» [en línia]. *JLIS.it: Italian Journal of Library, Archives, and Information Science*, v. 9, n. 1 (2018). <<https://www.jlis.it/article/view/12447>> [Consulta: 08/11/2018].

SULÉ DUESA, Andreu. «Schema.org: la millora de la visualització dels resultats en els cercadors i molt més» [en línia]. *BID: textos universitaris de biblioteconomia i documentació*, n. 34 (2015). <<http://dx.doi.org/10.1344/BID2015.34.23>> [Consulta: 08/11/2018].

SUMMERS, Ed, *et al.* «LCSH, SKOS and linked data» [en línia]. En: International Conference on Dublin Core and Metadata Applications (2008: Berlin). *DC-2008: proceedings of the International Conference on Dublin Core and Metadata Applications*. [Berlin: German National Library, 2008], p. 25-33. <<http://dcpapers.dublincore.org/pubs/article/view/916>> [Consulta: 08/11/2018].

TORRE BASTIDA, Ana Isabel; GONZÁLEZ RODRÍGUEZ, Marta; VILLAR RODRÍGUEZ, Esther. «Datos abiertos enlazados (LOD) y su implantación en bibliotecas: iniciativas y tecnologías» [en línia]. *El Profesional de la información*, v. 24, n. 2 (2015), p. 113-120. <<https://doi.org/10.3145/epi.2015.mar.04>> [Consulta: 08/11/2018]. ■



**Rubén ALCARAZ
MARTÍNEZ**

Departament de Biblioteconomia,
Documentació i Comunicació
Audiovisual, Universitat de
Barcelona
ralcaraz@ub.edu

Article rebut el juny de 2018;
revisat el juliol de 2018.

Les Google Accelerated Mobile Pages: la velocitat de càrrega com a factor de posicionament SEO

Resum: Es presenta la iniciativa Accelerated Mobile Pages de Google, se'n mostren els principals components i es discuteixen els beneficis que aporta com a tecnologia pensada per a millorar la velocitat dels llocs web per a dispositius mòbils i l'impacte que té en el posicionament dels resultats orgànics del cercador Google.

Paraules clau: Accelerated Mobile Pages, web mòbil, disseny web responsiu, optimització natural, SEO, WPO.

Las Google Accelerated Mobile Pages: la velocidad de carga como factor de posicionamiento SEO

Resumen: Se presenta la iniciativa Accelerated Mobile Pages de Google, se muestran sus principales Componentes y se discuten sus beneficios como tecnología pensada para mejorar la velocidad de los sitios web para dispositivos móviles y su impacto en el posicionamiento en los resultados orgánicos del buscador Google.

Palabras clave: Accelerated Mobile Pages, web móvil, diseño web responsivo, optimización natural, SEO, WPO.

Google Accelerated Mobile Pages: page speed as an SEO ranking factor

Abstract: Google's Accelerated Mobile Pages initiative is presented, its main components and its benefits as a technology designed to improve the speed of websites for mobile devices and their impact on the positioning of the organic results of the Google search engine are outlined and discussed.

Keywords: Accelerated Mobile Pages, mobile web, responsive web design, natural referencing, SEO, WPO.

1. Introducció

En els darrers anys, els dispositius mòbils, i específicament els telèfons intel·ligents, s'han consolidat com el principal mitjà a través del qual accedim al World Wide Web. L'any 2018 els telèfons intel·ligents van arribar a més de cinc mil milions de persones, això suposa una penetració del 68 %.¹

En el cas d'Espanya, que sempre és capdavantera en penetració de tecnologia mòbil al món, i segons la font consultada, entre un 80 %² i un 88 %³ dels espanyols són propietaris d'un telèfon intel·ligent. D'aquests, entre un 85 %⁴ i un 94,6 % l'utilitzen per accedir a internet. Segons l'*Encuesta sobre equipamiento y uso de tecnologías de información y comunicación en los hogares de 2017*, el 90,4 % dels espanyols van accedir a internet a través d'un dispositiu mòbil fora de la llar o el centre de treball. Les dades del Google Consumer Barometer són similars, amb un 91 % dels enquestats espanyols.

Pel que fa a les activitats dutes a terme des de dispositius mòbils, el 68 % dels espanyols accedeixen a les xarxes socials, el 79 % fan servir motors de cerca i el 64 % visualitzen vídeos amb reproducció en línia (*streaming*), una activitat que ha crescut per la disponibilitat de més bones connexions a la xarxa. Aquestes tres activitats, destacades per Google en el seu informe com a «comportaments clau», han registrat un augment en el nombre d'usuaris totals (fix i mòbil) gràcies a la ubicïtat pròpia dels dispositius mòbils.

1. *Digital in 2018: essential insights into Internet, social media, mobile, and ecommerce use around the world* [en línia], [New York: Hootsuite, 2018], <<https://hootsuite.com/uploads/images/stock/Digital-in-2018-001-Global-Overview-Report-v1.02-L.pdf>> [Consulta: 29/05/2018].
2. *La tecnologia digital en España en 2018* [en línia], [New York: Hootsuite, 2018], <https://hootsuite-online-revenue.s3.amazonaws.com/Digital_in_2018_Local_country_report/DIGITAL_IN_2018_006_SPAIN_v1.01.pdf> [Consulta: 29/05/2018].
3. *Informe mobile en España y en el mundo 2017* [en línia], Madrid: Ditrendia, 2017, <https://www.amic.media/media/files/file_352_1289.pdf> [Consulta: 2/06/2018].

Respecte a l'ús de cercadors, que són el principal punt d'accés al contingut disponible a la xarxa des de fa més d'una dècada, el percentatge de consultes dels usuaris des de dispositius mòbils augmenta any rere any, fins a superar les consultes que es fan des d'ordinadors de taula en països com Espanya (47 % mòbils i 8 % tauletes), els Estats Units (58 % mòbils i 7 % tauletes), el Regne Unit (46 % mòbils i 14 % tauletes), Itàlia (50 % mòbils i 7 % tauletes), Brasil (55 % mòbils i 2 % tauletes) i l'Índia (66 % mòbils i 2 % tauletes), una tendència a l'alça que ha implicat un 6 % d'augment d'usuaris el darrer any (Terenteva, 2017).

En el cas de les notícies digitals, la suma de consultes de mòbils i tauletes superen, d'ençà del 2017, l'accés a través d'ordinadors de taula (47 % enfront del 46 %).⁵ Un comportament que va créixer del 32 %, el 2014, fins al 47 % el 2017, i que és equiparable al creixement de l'ús de motors de cer-

Tant la indústria dels cercadors, amb Google al capdavant, com la dels editors de contingut i publicitat, lideren diferents projectes per gaudir d'una posició preferent en l'àmbit mòbil.

ca o botigues electròniques des dels dispositius mòbils. Aquesta és una tendència protagonitzada pels menors de 45 anys, dels quals el 61 % s'informa principalment a través dels mòbils (51 %) i, en menor mesura, de les tauletes (10 %).

Aquest context tecnològic, caracteritzat per l'increment en l'ús dels mòbils com a dispositius preferents per accedir a la informació, ha implicat l'aparició de diferents tecnologies per

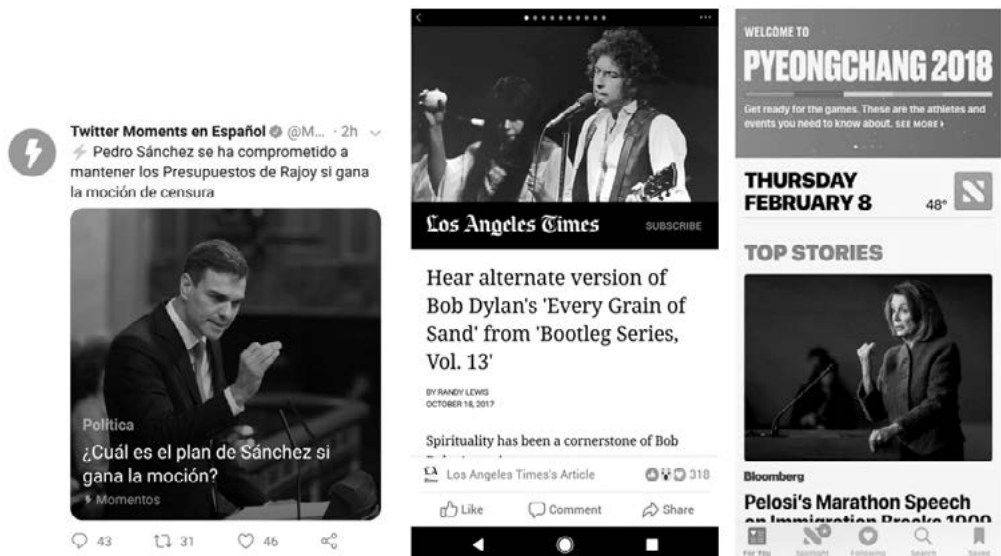


Figura 1. Twitter Moments, Facebook Instant Articles i Apple News són la resposta de Facebook, Twitter i Apple al nou escenari mòbil.

4. *Estudio de Google Consumer Barometer 2017: el año de los móviles* [en línia] (Dic. 2017), <<https://www.thinkwithgoogle.com/intl/es-es/canales-de-publicidad/movil/estudio-de-google-consumer-barometer-2017-el-%C3%B1o-de-los-m%C3%B3viles/>> [Consulta: 31/05/2018].
5. Alfonso Vara Miguel, Samuel Negrodo, Avelino Amoedo, *Digital news report.es 2017* [en línia]. Pamplona: Universidad de Navarra, Facultad de Comunicación, Center for Internet Studies and Digital Life, 2017, <<https://drive.google.com/file/d/0B2eyawMqcpTyb2FDZUNUoJmalEw>> [Consulta: 30/05/2018].

part dels principals distribuïdors de contingut a internet. En aquest sentit, tant la indústria dels cercadors, amb Google al capdavant, com la dels editors de contingut i publicitat, lideren diferents projectes per gaudir d'una posició preferent en l'àmbit mòbil. Les Accelerated Mobile Pages (AMP) de Google que citem en aquest article en són un cas destacat, però no l'únic. Plataformes com ara Facebook, Twitter i Apple també han entrat en aquest mercat amb els projectes Facebook Instant Articles,⁶ Twitter Moments⁷ i Apple News,⁸ respectivament.⁹

1.1. Del «Mobilegeddon» al Mobile-First Index

Google no ha passat per alt el context que hem descrit i, en els darrers anys, ha incorporat canvis substancials al seu algorisme amb l'objectiu d'adaptar-se a la realitat mòbil que caracteritza l'actual accés a internet.

La primera gran actualització de l'algorisme va arribar l'abril de 2015, amb l'anomenat «*Mobilegeddon*»,¹⁰ per incorporar com a factor de posicionament l'optimització dels llocs web per a dispositius mòbils. És a dir, els webs optimitzats per a dispositius mòbils havien de gaudir de més bon posicionament que aquells que no ho estiguessin. El fet que Google anunciés aquest canvi en l'algorisme dos mesos abans d'implementar-lo¹¹ i¹² i que manifestés que això tindria un «impacte significatiu», va alertar els

professionals del SEO i els desenvolupadors de tot el món. D'aquesta manera, Google volia alinear el comportament del seu cercador amb el comportament d'un usuari amb les tendències d'un mercat en què el mòbil era cada vegada més protagonista.

La realitat és que el nou factor no va tenir l'impacte previst. Tot i que es van observar canvis en els rànquings de resultats de les cerques des dels dispositius mòbils, van ser molt reduïts.¹³ Alguns experts van determinar que altres factors, com la velocitat de càrrega dels llocs web, podien haver tingut un impacte més fort que no pas el disseny optimitzat per a mòbils. La veritat és que Google va aconseguir, en només dos mesos, que milions de llocs web es redissenessin per a funcionar correctament en els dispositius mòbils, gràcies a l'anhelada recompensa de pujar posicions —o de no perdre-les— en el rànquing. Amb tot, la resposta de la indústria no va ser tan negativa com en el cas d'altres canvis en l'algorisme, segurament perquè en aquesta ocasió, a diferència d'altres casos anteriors, es va donar prou temps i les indicacions necessàries als professionals per adaptar-se als canvis.

Entre els mesos de març i maig de 2016, Google implementa un nou canvi relacionat amb la cerca mòbil al seu algorisme, conegut com el «*Mobilegeddon 2.0*».¹⁴ Aquesta nova actualització cercava potenciar el senyal inclòs en l'algorisme de l'any anterior per beneficiar, encara més, els llocs web optimitzats per a dispositius mòbils en les cerques mòbils.

6. <<https://instantarticles.fb.com>>.

7. <<https://twitter.com/i/moments>>.

8. <<https://www.apple.com/news>>.

9. Javier Guallar, «Prensa digital 2015-2017: los medios frente a las plataformas tecnológicas» [en línia], *Anuario ThinkEPI*, v. 12 (2018), <<https://recyt.fecyt.es/index.php/ThinkEPI/article/view/thinkepi.2018.33>> [Consulta: 29/05/2018].

10. També conegut com el *mobilepocalypse*, *mopocalypse* i *mobocalypse*.

11. Takaki Makino, Jung Chaesang, Doantam Phan, «Finding more mobile-friendly search results» [en línia], *Google webmaster central blog: official news on crawling and indexing sites for the Google index* (Feb. 26, 2015), <<https://webmasters.googleblog.com/2015/02/finding-more-mobile-friendly-search.html>> [Consulta: 27/05/2018].

12. Google acostuma a ser poc transparent amb tot allò que envolta el seu algorisme i rarament anuncia canvis abans d'implementar-los.

13. Peter J. Meyers, «7 days after Mobilegeddon: how far did the sky fall?» [en línia], *Moz blog* (Apr. 22, 2015), <<https://moz.com/blog/day-after-mobilegeddon>> [Consulta: 27/05/2018].

14. Klemen Kloboves, «Continuing to make the web more mobile friendly» [en línia], *Google webmaster central blog: official news on crawling and indexing sites for the Google index* (March 16, 2016), <<https://webmasters.googleblog.com/2016/03/continuing-to-make-web-more-mobile.html>> [Consulta 27/05/2018].

La darrera gran actualització de l'algorisme va arribar el març de 2018 amb el *mobile-first index*.¹⁵ Fins a aquesta actualització, el sistema de rastreig i indexació dels llocs web es basava en les versions per a ordinador de taula per a calcular la rellevància del resultat per a les consultes dels usuaris. En un context en el qual els usuaris ja fem més cerques des dels dispositius mòbils que des dels ordinadors de taula, aquesta manera de funcionar perd sentit, fins al punt que pot significar que els usuaris rebien resultats molt menys rellevants que si es tingués en compte de manera preferent el contingut de les versions mòbils dels llocs web. L'actualització *mobile-first index* implica que Google fa servir la versió mòbil de les pàgines per indexar-ne el contingut. Un únic índex en el qual Google ha començat a donar preferència al contingut que es distribueix als dispositius mòbils.

1.2. La velocitat de càrrega i l'impacte que té en les conversions i l'experiència d'usuari

Tothom ha experimentat en algun moment la frustració que representa accedir a un contingut disponible a internet i que no acabi de veure's mai. Una situació que encara empitjora més quan ens aborda la publicitat intrusiva present en molts llocs web. Els resultats d'un estudi de la companyia multinacional sueca Ericsson,¹⁶ en el qual es va mesurar la càrrega cognitiva¹⁷ dels participants davant de diferents situacions estressants, va concloure que el ritme cardíac dels participants s'elevava un 38 % i que el seu nivell d'estrès pujava considerablement en experimentar retards importants en el temps de càrrega d'una pàgina web, fins al punt que es tracta d'una situació comparable a l'estrès generat en veure una pel·lícula de terror o resoldre un problema matemàtic.

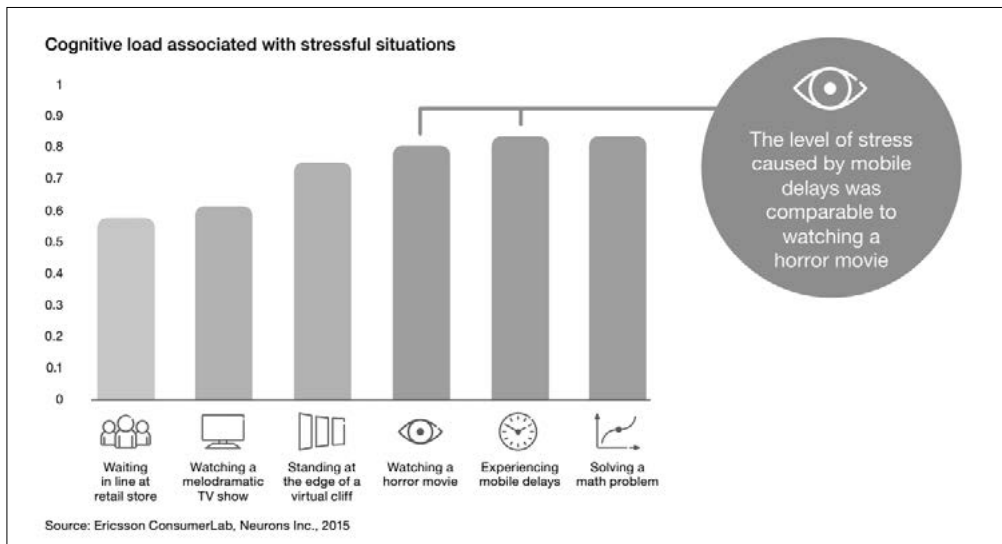


Figura 2. Càrrega cognitiva associada a diferents situacions estressants.

15. Fan Zhang, «Rolling out mobile-first indexing» [en línia], *Google webmaster central blog: official news on crawling and indexing sites for the Google index* (March 26, 2018), <<https://webmasters.googleblog.com/2018/03/rolling-out-mobile-first-indexing.html>> [Consulta: 27/05/2018].
16. *Ericsson mobility report: on the pulse of the networked society* [en línia], Stockholm: Ericsson AB, 2016, <<https://www.ericsson.com/assets/local/mobility-report/documents/2016/ericsson-mobility-report-feb-2016-interim.pdf>> [Consulta: 26/05/2018].
17. En el context de l'estudi d'Ericsson, s'entén per càrrega cognitiva una puntuació que mesura la quantitat d'informació que una persona processa mantenint activa l'escorça prefrontal dors lateral (una part del cervell). Les respostes davant de situacions estressants es caracteritzen per presentar càrregues cognitives elevades. Es considera acceptable una puntuació de fins a 0,7, mentre que puntuacions de 0,8 o superiors es consideren molt altes i estressants.

El problema de la velocitat de càrrega, terme conegut en la literatura especialitzada com el rendiment (*performance*) dels llocs web, ha preocupat en les darreres dues dècades a les principals companyies d'internet, que han creat departaments específics per a millorar el rendiment dels seus productes. El conjunt de tècniques relacionades amb la millora del rendiment dels llocs web es coneix amb el nom de Web Performance Optimization (WPO) i han arribat al públic especialitzat a partir dels responsables dels departaments d'optimització del rendiment de companyies com Google i Yahoo!. Així mateix, Killelea,¹⁸ Souders,^{19 i 20} i Zakas²¹ i Frick²² han publicat diversos manuals especialitzats en la matèria.

La velocitat de càrrega de cadascuna de les pàgines d'un lloc web s'acostuma a expressar en els mil·lisegons o segons que triga a carregar-se una pàgina en un context de laboratori. Per Walton,²³ les afirmacions del tipus «he calculat el temps de càrrega de la meua aplicació —o lloc web— i és de X,XX segons» són falses,

ja que el temps de càrrega d'una pàgina web pot variar dràsticament segons les característiques del dispositiu des del qual es fa la consulta o el tipus de connexió disponible.

Un altre aspecte que cal considerar és la dificultat de determinar quan s'ha carregat la pàgina. En general, s'acostuma a pensar que la pàgina es carrega quan tots els recursos necessaris perquè es visualitzi i funcioni correctament (fitxers HTML, fulls d'estil, fitxers JavaScript, imatges, etc.) s'han descarregat en l'ordinador de l'usuari. Per Walton (2018), la càrrega no arriba en un moment concret, sinó que es tracta d'una experiència completa. En alguns casos, quan accedim a un lloc web no podem veure el contingut fins que la càrrega no s'ha completat del tot, mentre que, en d'altres, el contingut es mostra des d'un primer moment —almenys parcialment—, tot i que algunes funcionalitats dependents del JavaScript de la pàgina encara no funcionin. En el primer cas, la sensació de lentitud és molt més acusada, encara que el temps de càrrega total d'ambdues sigui el mateix per a l'usuari.

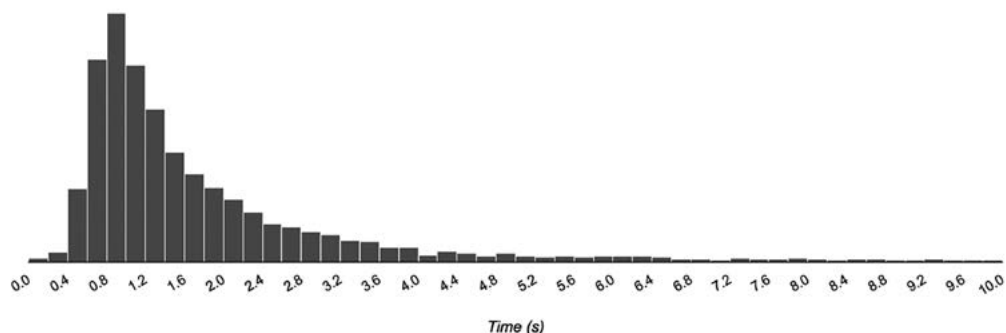


Figura 3. Walton²⁴ entén el temps de càrrega com la col·lecció de tots els temps de càrrega de cadascun dels usuaris que han consultat la pàgina o aplicació. En aquest histograma podem observar com la moda se situa entorn d'un segon, però per alguns altres usuaris el lloc o aplicació web carrega en només 0,4 mil·lisegons, i per altres pot arribar a trigar 10 vegades més. Per tant, el temps de càrrega s'ha d'entendre com una distribució i no com un valor absolut.

18. Patrick Killelea, *Web performance tuning*, Sebastopol, CA.: O'Reilly, cop. 2002.

19. Steve Souders, *High performance web sites: essential knowledge for frontend engineers*, Sebastopol, CA [etc.]: O'Reilly, 2007.

20. Steve Souders, *Even faster websites*, Sebastopol, CA: O'Reilly, 2009.

21. Nicholas C. Zakas, *High performance JavaScript: build faster web application interfaces*, Sebastopol, CA: O'Reilly; Yahoo! Press, 2010.

22. Tim Frick, *Designing for sustainability: a guide to building greener digital products and services*. Sebastopol, CA: O'Reilly, 2016.

23. Philip Walton, «User-centric performance metrics» [en línia], *Web fundamentals: Google's opinionated reference for building amazing web experiences* (2018), <<https://developers.google.com/web/fundamentals/performance/user-centric-performance-metrics>> [Consulta: 26/05/2018].

24. Philip Walton, *op cit.*



Figura 4. Procés de càrrega d'un lloc web fotograma a fotograma amb indicació de les diferents fites assolides.²⁷

Entre les analítiques web de rendiment centrades en l'usuari que Google fa servir per determinar el rendiment d'un lloc web, en destaquen el First Contentful Paint (FCP) i el DOMContentLoaded (DCL). La primera fa referència al moment en el qual els usuaris obtenen la primera resposta visual de la pàgina. Aquest moment és important perquè com més aviat arribi més possibilitats hi ha de mantenir l'interès de l'usuari. D'altra banda, el DCL mesura el moment en el qual es carrega tot el document HTML. Per assignar un valor d'FCP i DCL a cada pàgina web, així com d'altres mètriques relacionades amb la velocitat, Google utilitza el CrUX (Chrome User Experience Report),²⁵ una eina disponible dins del servei de pagament Google BigQuery²⁶ que es nodreix de les dades que els usuaris del navegador Google Chrome comparteixen amb Google.

Tant l'FCP com el DCL es poden consultar per URL específics mitjançant l'eina Google PageSpeed Insights²⁸ i el Lighthouse,²⁹ una eina de codi obert de Google pensada per millorar la qualitat de les aplicacions web.

Però quin és l'impacte real de la velocitat de càrrega en aspectes com l'experiència d'usuari

i la capacitat dels llocs web d'assolir els seus objectius? Segons dades de Google,³⁰ a principi del 2017, més del 75 % de les versions mòbils dels llocs web trigaven més de deu segons a carregar, i la mitjana de càrrega de les pàgines d'inici era de dinou segons en xarxes 3G i catorze en xarxes 4G. Mentrestant, el 53 % dels internautes abandonaven els llocs web després dels tres primers segons de càrrega.

Des del punt de vista dels ingressos per publicitat, els llocs web mòbils que carregaven en menys de cinc segons obtenien el doble d'ingressos que els que ho feien en dinou segons. No obstant això, la necessitat de monetitzar el contingut com a principal estratègia dels llocs web de contingut (premsa, blogs, etc.) provoca, en molts casos, llocs sobrecarregats de publicitat extremadament lents i que ofereixen una molt mala experiència d'usuari.

1.3. La velocitat de càrrega com a factor de posicionament

A la capacitat d'un lloc web per monetitzar el contingut i a la necessitat d'oferir una bona experiència d'usuari, cal afegir-hi l'impacte que té

25. <<https://developers.google.com/web/tools/chrome-user-experience-report>>.

26. <<https://cloud.google.com/bigquery>>.

27. Philip Walton, *op. cit.*

28. <<https://developers.google.com/speed/pagespeed/insights/?hl=ca>>.

29. <<https://developers.google.com/web/tools/lighthouse>>.

30. «The need for mobile speed: how mobile latency impacts publisher revenue» [en línia], *Think with Google* (Sep. 2016), <<https://www.thinkwithgoogle.com/intl/en-154/insights-inspiration/research-data/need-mobile-speed-how-mobile-latency-impacts-publisher-revenue/>> [Consulta: 29/05/2018].

el temps de càrrega de les pàgines com a factor de posicionament. És a dir, en el càlcul de la rellevància d'un lloc web per una consulta determinada el cercador de Google té en compte, entre moltíssims altres factors, la velocitat de càrrega de cadascuna de les pàgines que té al seu índex, i les ofereix a l'usuari com a resultats a la Search Engine Results Page (SERP).

L'any 2010 Singhal i Cutts,³¹ dos enginyers del Google Search Quality Team, informaven, en el blog Google Webmaster Central blog, el canal oficial de Google sobre qüestions relacionades amb el seu cercador, de la incorporació a l'algorisme de la velocitat de càrrega de les pàgines com a factor de posicionament per a les cerques realitzades des d'ordinadors de taula. El gener de 2018 Wang i Phan³² oficialitzaven la incorporació del mateix factor al rànquing mòbil a partir del mes de juliol del mateix any.

A part de la incorporació a l'algorisme com a factor de posicionament, la velocitat de càrrega de les pàgines d'un lloc web també pot ser un condicionant important per a indexar llocs web especialment extensos (més d'un miler de pàgines). En aquest sentit, Google ha reconegut que assigna un «pressupost de rastreig» (*crawl budget*) al Googlebot, l'eina encarregada de rastrejar la xarxa, localitzar noves pàgines i incorporar-les al seu índex. Aquest pressupost es determina a partir de diferents paràmetres, entre els quals intervenen diferents aspectes, com les velocitats de càrrega, senyal que indica, segons Illyes,³³ la bona salut del servidor en què s'allotja el lloc web i la possibilitat per al Googlebot de descobrir més contingut amb el mateix nombre de connexions.³⁴

2. Les Accelerated Mobile Pages

2.1. Definició i motivació

Les Accelerated Mobile Pages (AMP) són una iniciativa de codi obert de Google que tenen l'objectiu de permetre als desenvolupadors crear pàgines web i publicitat molt més ràpides, amb independència del dispositiu o la plataforma a través del qual es distribueixi aquest contingut.

El projecte AMP es va estrenar l'any 2015, en vista de la dificultat dels distribuïdors de continguts d'implementar tècniques per optimitzar el rendiment dels llocs web.

D'aquesta manera, Google volia apropar el Web Performance Optimization a tots els desenvolupadors, ja fossin un gran mitjà de comunicació amb importants recursos o un petit actor local.

2.2. Components

Les AMP es basen en tres components principals: AMP HTML, AMP JS i AMP Cache.

- AMP HTML és un subconjunt restringit de l'estàndard HTML amb algunes limitacions en l'ús de diverses etiquetes.
- AMP JS és una biblioteca de JavaScript que permet renderitzar d'una manera molt més ràpida les pàgines creades amb AMP HTML.

31. Amit Singhal, Matt Cutts, «Using site speed in web search ranking» [en línia], *Google webmaster central blog: official news on crawling and indexing sites for the Google index* (Apr. 9, 2010), <<https://webmasters.googleblog.com/2010/04/using-site-speed-in-web-search-ranking.html>> [Consulta: 25/05/2018].

32. Zhiheng Wang, Doantam Phan, «Using page speed in mobile search ranking» [en línia], *Google webmaster central blog: official news on crawling and indexing sites for the Google index* (Jan. 17, 2018), <<https://webmasters.googleblog.com/2018/01/using-page-speed-in-mobile-search.html>> [Consulta: 25/05/2018].

33. Gary Illyes, «What crawl budget means for Googlebot» [en línia]. *Google webmaster central blog: official news on crawling and indexing sites for the Google index* (Jan. 16, 2017), <<https://webmasters.googleblog.com/2017/01/what-crawl-budget-means-for-googlebot.html>> [Consulta: 25/05/2018].

34. En aquest context, entenem per connexions cadascuna de les peticions que fa un usuari o un robot com el Googlebot quan descarrega una pàgina web. Aquestes inclouen tots i cadascun dels fitxers relacionats amb la pàgina, com ara els fulls d'estil, els fitxers JavaScript i les imatges, entre d'altres.

- AMP Cache és una tècnica emprada per a obtenir les pàgines AMP, guardar-les en una memòria cau i servir-les des d'aquesta. També s'encarrega de validar les pàgines amb l'objectiu de mostrar només pàgines correctament implementades.

2.2.1. AMP HTML

AMP HTML parteix de l'estàndard HTML5 i és totalment compatible amb els principals navegadors web del mercat.³⁵ Entre les etiquetes i els atributs propis d'aquest subconjunt, en podem diferenciar de tres tipus: obligatòries, prohibides —en alguns casos amb una alternativa específica— i opcionals.



Figura 5. A l'esquerra, pàgina web sense la finestra gràfica adaptada a la mida del dispositiu —en aquest cas, un mòbil. A la dreta, la mateixa pàgina amb la finestra gràfica configurada per adaptar-se a la mida de la pantalla.³⁶

La primera etiqueta obligatòria que trobem en qualsevol document AMP és la definició de tipus de document (`<!doctype html>`), una etiqueta que també és obligatòria en qualsevol document HTML, sigui AMP o no. A partir d'aquí, l'especificació AMP estableix com

Les Accelerated Mobile Pages (AMP) són una iniciativa de codi obert de Google que tenen l'objectiu de permetre als desenvolupadors crear pàgines web i publicitat molt més ràpides, amb independència del dispositiu o la plataforma a través del qual es distribueixi aquest contingut.

a obligatòries altres etiquetes que no ho són en l'estàndard del W3C, com el `<head>` i el `<body>`, la configuració de la codificació de caràcters, la de configuració de la finestra gràfica (*viewport*) del dispositiu o l'enllaç a la pàgina canònica amb el contingut original (fix o mòbil), del qual parlarem més endavant.

L'ús del JavaScript està totalment prohibit a les pàgines AMP, amb l'excepció de la biblioteca AMP JS i dels casos en els quals s'utilitza per incorporar dades en format JSON-LD.³⁷ A més d'aquesta restricció, trobem altres etiquetes de l'estàndard HTML5 no permeses, com: `<base>`, `<frame>`, `<frameset>`, `<object>`, `<param>`, `<applet>` o `<embed>`, algunes de les quals estan tradicionalment associades a un increment del temps de càrrega de les pàgines web.

En altres casos, l'etiqueta es prohibeix, però s'ofereix una alternativa AMP. Aquest és el cas d'etiquetes tan necessàries com l'etiqueta ``,³⁸ prohibida i substituïda per l'etiqueta `<amp-img>`. A part de les imatges, altres tipus de continguts com els vídeos (`<amp-video>`), els àudios (`<amp-audio>`) i els *iframes* (`<amp-iframe>`) també tenen una etiqueta específica per a optimitzar-ne el temps de càrrega.

35. AMP és compatible amb les darreres versions dels navegadors Chrome, Firefox, Edge, Safari, Opera i UC Browser.

36. `<https://www.w3schools.com/css/css_rwd_viewport.asp>`.

37. JSON-LD (JavaScript Object Notation for Linked Data) és l'estàndard adoptat per l'W3C per a la codificació i transmissió de dades enllaçades. Està basat en l'estàndard per a l'intercanvi de dades JSON.

38. En l'estàndard HTML, l'etiqueta `` s'utilitza per afegir imatges a les pàgines web.

La resta d'etiquetes de l'estàndard HTML5, com en les etiquetes d'estructura semàntiques (*article, section, aside, footer...*), els diferents nivells d'encapçalaments (h1, h2, h3...) són opcionals.

Com qualsevol altra pàgina web, les pàgines AMP fan servir l'estàndard CSS per donar estil al document. La diferència entre una pàgina HTML corrent i una AMP és que la darrera no admet fulls d'estil externs, una pràctica habitual en la immensa majoria de llocs web. El principal avantatge dels fulls d'estil externs és la possibilitat de crear un únic full amb l'estil de totes les pàgines del lloc web, en comptes de repetir-lo en totes. Tanmateix, aquesta pràctica implica la necessitat de carregar un recurs addicional —el fitxer CSS— que moltes vegades tampoc no s'optimitza. Per evitar aquest problema, AMP HTML proposa l'ús del CSS en línia, és a dir, dins del mateix fitxer HTML. L'única excepció la trobem en l'ús de fonts tipogràfiques personalitzades procedents d'una llista blanca de repositoris especialitzats.³⁹

Per acabar aquest apartat, cal destacar que l'AMP HTML inclou tot un conjunt de components com ara reproductors de vídeo específics, suport per incorporar etiquetes de xarxes socials o fórmules matemàtiques en format MathML, entre d'altres. Tots aquests continguts que generalment s'incorporen al codi HTML a partir de biblioteques específiques o essent incrustats directament en el codi font requereixen, perquè la pàgina AMP sigui correcta sintàcticament, l'ús de cada una de les extensions específiques desenvolupades per al projecte AMP.⁴⁰

2.2.2. AMP JS

El JavaScript (JS) és el llenguatge de programació per excel·lència de la xarxa. Gràcies al JavaScript és possible afegir dinamisme a les pàgines web, modificar-ne l'aspecte (contingut o estil), adaptar la manera d'interactuar dels usuaris amb el contingut, agilitzar la comunicació entre el client (navegador) i el servidor (màquina o es troben les aplicacions o llocs web), i validar dades abans d'enviar-les al servidor, entre d'altres. Tanmateix, un ús inadequat o excessiu del JavaScript pot implicar problemes de rendiment en els llocs web.

Amb l'objectiu d'evitar els problemes derivats d'un ús inadequat del JavaScript, Google prohibeix l'ús de qualsevol biblioteca o codi escrits en aquest llenguatge, que no siguin les pròpies del projecte AMP.

La biblioteca AMP JS va néixer amb l'objectiu de donar suport a l'ús de les etiquetes AMP HTML no estàndards, així com per optimitzar al màxim la càrrega del contingut. En resum, l'AMP JS es caracteritza per:

- Permetre només l'ús del JavaScript asíncron.⁴¹
- Evitar que determinats recursos com, per exemple, el codi incrustat per mostrar tuits o fotografies d'Instagram bloquegin la representació de la pàgina fins que la sol·licitud HTTP necessària per importar aquest recurs hagi finalitzat.
- Prioritzar la càrrega dels recursos més importants.
- Optimitzar l'activació de fonts tipogràfiques importades des de repositoris acceptats.
- Optimitzar la càrrega dels estils CSS en línia.

39. Entre els llocs permesos trobem Google Fonts, Typography.com, Fonts.com, Typekit i Font Awesome.

40. La llista completa de components es pot consultar a: <<https://www.ampproject.org/docs/reference/components>>.

41. Quan un usuari arriba a una pàgina web espera la resposta del servidor fins que el codi HTML arriba al navegador i mostra la pàgina sol·licitada. Aquest és un comportament de comunicació síncrona, en el qual el client (navegador) ha enviat una petició i roman bloquejat esperant la resposta del servidor. Els mecanismes de comunicació asíncrona permeten recarregar en segon pla una part de la pàgina web, deixant desbloquejada la resta. Això permet que les aplicacions i els llocs web tinguin una interactivitat similar a la de les aplicacions per a ordinador de taula. Alguns exemples de comunicació asíncrona els trobem a la safata d'entrada del Gmail, o en els marcadors dels diaris esportius digitals que s'actualitzen sense que l'usuari hagi de recarregar la pàgina.

2.2.3. AMP Cache

Les pàgines AMP es guarden al mateix servidor que la resta de pàgines del lloc web. Tanmateix, estan pensades perquè tercers actors les obtinguin, validin i, si la seva implementació és correcta, les desin en una memòria cau específica.

L'AMP Cache és una xarxa de lliurament de contingut basada en servidors intermediaris (*proxy-based CDN*) creada per a distribuir pàgines AMP vàlides. Aquesta infraestructura mundial propietat de Google permet als editors de contingut gaudir dels avantatges d'una arquitectura basada en la replicació de versions del contingut en diferents servidors distribuïts per tot el món, els quals donen servei als usuaris d'acord amb la seva proximitat geogràfica.

El sistema funciona de la manera següent:

- L'usuari fa una cerca a Google des d'un dispositiu mòbil.
- Google li mostra una pàgina de resultats de cerca (SERP) amb alguns resultats que es corresponen amb pàgines AMP validades i disponibles als servidors de Google.
- L'usuari fa clic sobre un resultat AMP.
- Google respon la petició de l'usuari enviant la pàgina sol·licitada des del servidor AMP ubicat geogràficament més a prop de l'usuari.

Es tracta d'una arquitectura complexa d'implementar i, sobretot, molt cara, que Google posa a l'abast de tots els desenvolupadors i editors que participen en el projecte AMP. A banda de millorar la velocitat amb optimitzacions addicionals al rendiment de les pàgines, l'AMP Cache té altres objectius, com ara assegurar que totes les pàgines disponibles siguin vàlides i no continguin errors de codi.⁴²

Quan l'usuari accedeix a una pàgina AMP allotjada en un dels servidors de Google, l'URL que visualitza en el seu navegador presenta l'estructura següent:

<https://www.google.es/amp/s/url-de-l'editor.html>

Aquest contingut no es troba allotjat al servidor del proveïdor de contingut, sinó al servidor AMP de Google més proper a la ubicació geogràfica des de la qual l'usuari ha fet la consulta.

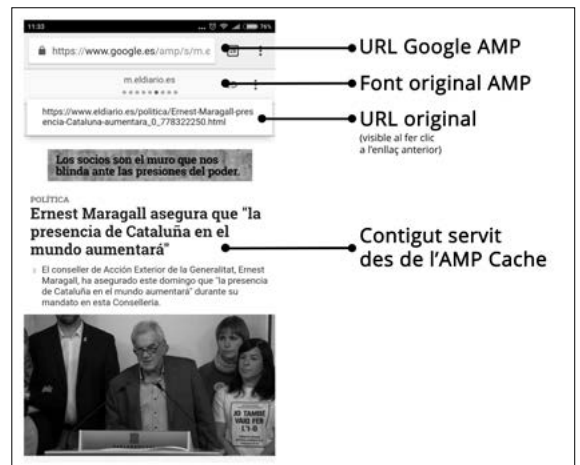


Figura 6. Parts d'una pàgina AMP disponible als servidors de Google.

Tot i que l'adreça on es troba la pàgina AMP que visita l'usuari no és la del lloc web propietari del contingut, sinó una adreça de Google, aquest atribueix tot el trànsit a l'editor mitjançant l'etiqueta <amp-analytics>. Això és important, no només pel pes que té el trànsit web com a factor d'optimització natural (SEO), sinó també per poder monetitzar la publicitat incrustada des de plataformes publicitàries com AdSense.

42. Paul Bakaus, «Why AMP Cache exist» [en línia], *Medium* (2017), <<https://medium.com/@pbakaus/why-amp-caches-exist-cd7938da2456>> [Consulta: 02/06/2018].
 43. Lucía Jiménez Iglesias, *et al.*, «Experiencia de usuario y medios de comunicación: la regla de los tres clics en las webs de periódicos para smartphones» [en línia], *Revista latina de comunicación social*, n. 73 (2018), p. 595-613, <<http://www.revistalatinacs.org/073paper/1271/30es.html>> [Consulta: 02/06/2018].

2.2.4. AMP HTML ADS

La publicitat a internet, especialment en el cas dels llocs web de contingut com els de la premsa digital o els blogs, és un dels aspectes que més insatisfacció causa en els usuaris. L'estudi de Jiménez, *et al.*⁴³ mostra opinions significatives en aquest sentit, en les quals es pot observar que els usuaris relacionen directament l'ús de la publicitat amb més dificultat per navegar entre continguts web, així com en la percepció del rendiment d'aquestes pàgines.

Per Fessenden⁴⁴ la publicitat a la xarxa ha de complir cinc requisits per garantir una bona experiència d'usuari: control de l'usuari (facilitar l'acció de tancar-la i evitar reproduccions o obertures automàtiques), gratificació instantània (evitar anuncis que dificultin o retardin l'accés al contingut principal), ubicació (evitar anuncis ubicats sobre el contingut o que dificultin interactuar-hi), predictibilitat (capacitat de l'usuari per entendre el comportament de l'anunci) i rellevància (que estigui relacionada

amb el contingut o els objectius del lloc web). La realitat, però, és que la publicitat a internet és lenta, impredecible i disruptiva.

En vista d'aquesta realitat, el projecte AMP preveu la incorporació d'anuncis en les seves pàgines que tenen en compte tant l'experiència de l'usuari com la necessitat d'obtenir ingressos dels editors. Aquests anuncis funcionen d'acord amb les especificacions AMP HTML, això vol dir que ja no se'ls permet utilitzar indiscriminadament el JavaScript, motiu principal pel qual els anuncis que tradicionalment trobem incrustats als llocs web són tan lents.

Entre els beneficis dels AMP HTML Ads destaquen: la major velocitat de càrrega, ja que se sol·liciten els anuncis abans del procés de representació del contingut a la pàgina; més eficiència en aturar-se els anuncis animats quan aquests es troben en l'espai que està visualitzant l'usuari, o més seguretat, ja que han de passar obligatòriament per un procés de validació molt estricte.

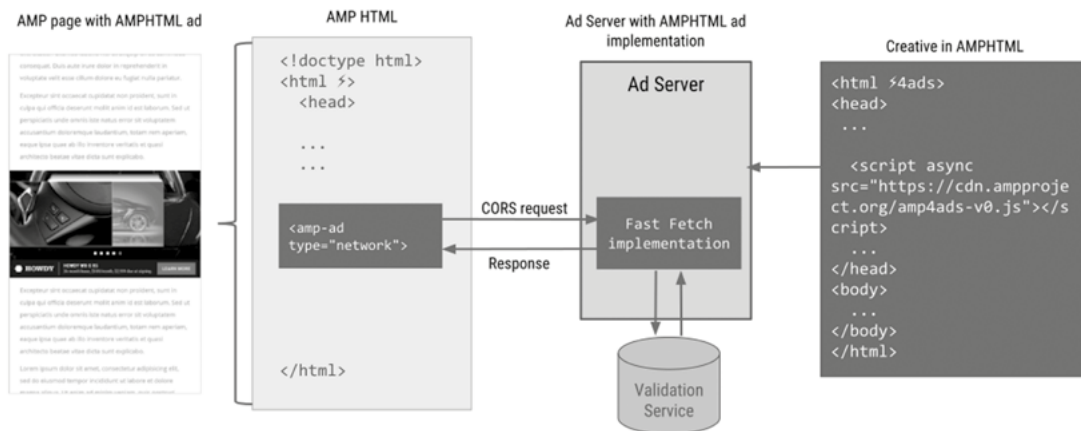


Figura 7. Funcionament dels AMP HTML ads.⁴⁵

44. Therese Fessenden, «Five user requirements for online ads» [en línia]. Nielsen Norman Group. Articles (July 16, 2017). <https://www.nngroup.com/articles/user-requirements-online-ads/> [Consulta: 05/06/2018].

45. <https://www.ampproject.org/es/docs/ads/amphtml_ads>.

2.3. Implementació

Google mai ha ocultat la seva preferència pel disseny web responsiu (*responsive web design*),⁴⁶ davant altres alternatives com ara les versions mòbils específiques i el disseny web adaptatiu (*adaptive web design*, també conegut com a *dynamic serving* o publicació dinàmica).

La raó és simple: per Google és més fàcil trobar i indexar el contingut d'un lloc web que es distribueix amb un mateix URL i que, tot i les adaptacions necessàries per funcionar d'una manera òptima en diferents dispositius, ofereix a tots els dispositius el mateix contingut. Però si Google es decanta pel disseny responsiu, quin és el sentit de les pàgines AMP? La resposta la trobem en el rendiment dels llocs web responsius quan es consulten des d'un mòbil i no s'han dissenyat des d'un punt de vista específic per a mòbils.

El disseny web responsiu es fonamenta en l'adaptació dels continguts d'una pàgina web a

qualsevol dispositiu o mida de pantalla a partir de diferents solucions tècniques que parteixen d'una única versió del codi font.⁴⁷ El principal avantatge del disseny responsiu és la possibilitat d'utilitzar un únic codi font per a totes les versions del lloc web, però aquest avantatge es converteix sovint en el seu principal problema, especialment quan es dissenya primer la versió per a ordinador de taula.

En dissenyar la versió per a ordinador de taula d'un lloc web responsiu, els desenvolupadors assumeixen que el seu lloc o aplicació mòbil funcionarà en un equip relativament potent, amb una pantalla gran, una bona connexió a la xarxa i que serà compatible amb totes les tecnologies de desenvolupament web més o menys estandarditzades. Els canvis en el disseny de la versió mòbil de la mateixa pàgina acostumen a limitar-se a adaptar la disposició dels elements a una retícula més simple, passant de l'ús de diverses columnes a una de sola, i a canvis en la mida dels elements de la pàgina (text, imatges, vídeos, etc.).

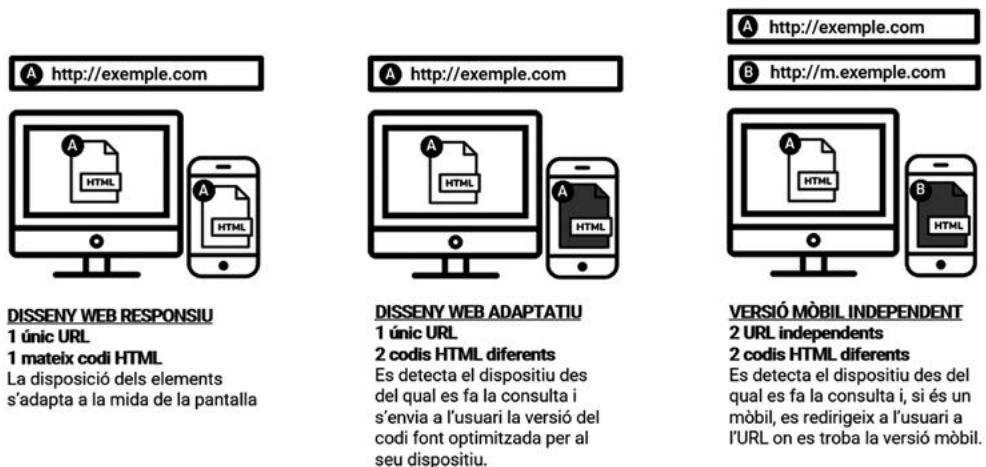


Figura 8. Característiques i trets diferencials del disseny web responsiu, el disseny adaptatiu i les versions mòbils independents.

46. Pierre Far, «Recommendations for building smartphone optimized websites [en línia]», *Google webmaster central blog: official news on crawling and indexing sites for the Google index* (June 2012), <<https://webmasters.googleblog.com/2012/06/recommendations-for-building-smartphone.html>> [Consulta: 29/05/2018].
47. Rubén Alcaraz Martínez, «Webs i aplicacions mòbils: una aproximació des del punt de vista de l'experiència d'usuari» [en línia], *Ítem: revista de biblioteconomia i documentació*, n. 61 (2016), p. 21-42, <<https://www.raco.cat/index.php/Item/article/view/317352>> [Consulta: 02/06/2018].

Aquesta metodologia, coneguda com a *desktop-first*, en contraposició a la *mobile-first*, generalment deriva en llocs web que, si bé s'adapten visualment a diferents mides de pantalla, ofereixen un rendiment molt pobre en els dispositius mòbils. Per una banda, la metodologia *mobile-first* advoca per dissenyar primer la versió mòbil, amb les limitacions que implica disposar de menys espai i funcionalitats, pensant primer en la millor experiència d'usuari per als visitants que accedeixen al contingut des d'un dispositiu mòbil i adaptant-ne, posteriorment, el contingut a pantalles més grans i dispositius més potents en què es poden oferir experiències d'usuari i interaccions més complexes.⁴⁸

La implementació de les pàgines AMP preveu i admet els diferents escenaris descrits a la figura 8. En aquest sentit, els llocs web que tenen un disseny web responsiu també poden oferir, en el cas de les consultes realitzades des de mòbils, accés a les versions AMP d'aquestes pàgines. La relació entre la versió «original» de la pàgina i la versió AMP es formalitza a través del mateix sistema de canonicalització que fem servir quan volem indicar a Google quin és l'URL preferit per a cada pàgina, en el cas que sigui possible accedir-hi mitjançant diferents adreces.⁴⁹ La relació s'estableix mitjançant les etiquetes següents, que s'han d'ubicar dins les capçaleres dels documents HTML «original» i de la versió AMP.

Etiqueta en la capçalera del document HTML «original»:
`<link rel="amphtml" href="URL a la versió AMP de la pàgina">`

Etiqueta en la capçalera de la versió AMP de la pàgina
`<link rel="canonical" href="URL on es troba la versió no AMP de la pàgina">`

En cas de disposar d'una versió per a ordinador de taula i d'una versió específica per a dispositius mòbils, s'afegeix una mica més de complexitat al procés, però el sistema és el mateix. En aquest cas, tant la versió mòbil com la versió AMP han de tenir un enllaç canònic a la versió per a ordinador de taula de la pàgina.

Etiqueta en la capçalera del document HTML «original»:

`<link rel="amphtml" href="URL a la versió AMP de la pàgina">`

Etiqueta en la capçalera de la versió mòbil de la pàgina:

`<link rel="canonical" href="URL on es troba la versió per a ordinador de taula de la pàgina">`

Etiqueta en la capçalera de la versió AMP de la pàgina:

`<link rel="canonical" href="URL on es troba la versió per a ordinador de taula de la pàgina">`

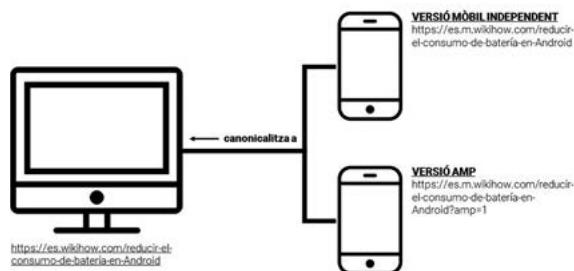


Figura 9. Funcionament del sistema de canonicalització proposat per Google.

En el cas dels llocs web que encara no s'han adaptat a l'ecosistema mòbil, o per a llocs web de nova creació, una altra possibilitat pot ser crear com a única versió mòbil del seu lloc les versions AMP de les seves pàgines.

48. Luke Wroblewski, *Mobile first*. New York, N.Y.: A Book Apart, 2011.

49. En alguns casos es pot accedir a una mateixa pàgina web mitjançant diferents URL. Per exemple, quan no s'ha implementat una redirecció entre la versió amb tres ves dobles i la versió sense, o com a conseqüència de la gestió interna dels URL que fan els sistemes de gestió de continguts. En aquests casos, convé indicar a Google quina de les diferents opcions és la preferida —o canònica.



Figura 10. D'esquerra a dreta, versió per a ordinador de taula, versió mòbil independent i versió AMP d'una pàgina del lloc web WikiHow. En la part superior de la versió AMP s'hi ha inclòs l'enllaç a la versió mòbil, de manera visible.

Aquesta gran flexibilitat en la implementació de les AMP, però, sobretot, el fet de no necessitar redissenyar tot el lloc web o la versió mòbil, sinó simplement d'afegir-les com una capa addicional que actua en el context mòbil, fa que la iniciativa de Google sigui molt atractiva per a molts desenvolupadors. És per aquest motiu que la creació de versions AMP és molt més assumible. També perquè es basa en l'ús de tècniques de desenvolupament conegudes i relativament fàcils d'implementar i, a més, compatibles amb alguns dels principals sistemes de gestió de continguts del mercat.⁵⁰

Un aspecte important que cal tenir en compte és que el contingut que es distribueix amb la versió AMP ha de ser el més similar possible al que es troba a la versió canònica de la mateixa pàgina. La versió AMP pot tenir un disseny més

simple i pot prescindir d'alguns elements o simplificar-los, com els menús de navegació, els peus de pàgina o les barres laterals que trobem tradicionalment als blogs amb informació sobre l'autor, enllaços a webs de tercers o publicitat. El que no pot variar de cap de les maneres és el contingut principal de la pàgina, és a dir, el títol i el cos. En el cas que hi hagi una diferència significativa entre el contingut principal d'una pàgina AMP i una no-AMP, Google redirigeix automàticament l'usuari cap a la versió no-AMP, entenent que aquesta té el contingut original i complet. Aquest criteri, lluny de ser una penalització — no implica cap mena de modificació en el rànquing —, sí que pot suposar una mala experiència d'usuari per als visitants que arriben a la versió no-AMP des dels seus dispositius mòbils, a més de la impossibilitat d'aparèixer al carrusel de notícies destacades.⁵¹

50. En la documentació de la iniciativa AMP trobem referències a les extensions disponibles per als principals sistemes de gestió de continguts del mercat: <https://www.ampproject.org/docs/getting_started/quickstart#create-your-amp-pages>.

51. Ashish Mehta, «Engaging users through high quality AMP pages» [en línia], *Google webmaster central blog: official news on crawling and indexing sites for the Google index* (Nov. 16, 2017). <<https://webmasters.googleblog.com/2017/11/engaging-users-through-high-quality-amp.html>> [Consulta: 29/05/2018].



2.4. Validació

La validació de les pàgines AMP és imprescindible per assegurar-ne la correcta implementació. Un sol error en el codi impedirà que aquestes pàgines apareguin en l'AMP Cache de Google o en llocs web de tercers que també han començat a mostrar contingut AMP, com ara Pinterest.

Per validar el codi font d'una pàgina AMP, Google ofereix diferents possibilitats. La primera és la més fàcil d'utilitzar pels usuaris amb menys coneixements tècnics. Es tracta d'un validador web⁵² que mostra, entre línies i en forma de resum a continuació del codi font, els errors detectats. A més de mostrar els errors, proporciona diferents recomanacions per resoldre'ls. Per fer-lo servir només cal indicar l'URL que volem validar i el tipus de pàgina AMP (genèrica, amb publicitat o per correu electrònic).⁵³

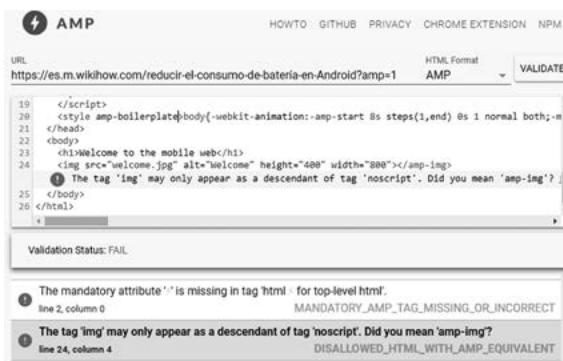


Figura 11. Fragment del resultat de la validació d'una pàgina AMP amb dos errors de codi font.

Una segona possibilitat és l'ús de l'extensió AMP validator disponible per als navegadors Chrome i Opera. A més d'aquestes dues opcions, també es pot validar el codi des de la consola de desenvolupadors integrada dins de Google Chrome,⁵⁴ o des de la interfície de línia d'ordres amb les extensions específiques disponibles per NPM.⁵⁵

2.5. Errors comuns en la implementació

Tot i l'extensa i completa documentació disponible, la majoria de les pàgines web en les quals s'ha implementat la tecnologia AMP presenten errors d'implementació. Segons un estudi de la consultora SEMrush,⁵⁶ el 75 % dels llocs web que les fan servir contenen, almenys, un error. En el cas d'Espanya, el 78,5 %.

La major part d'aquests errors estan relacionats amb l'ús d'etiquetes o atributs HTML no permesos o amb la manca d'alguns elements obligatoris. En segon lloc, l'estudi també mostra una relativa presència d'errors relacionats amb l'especificació CSS.

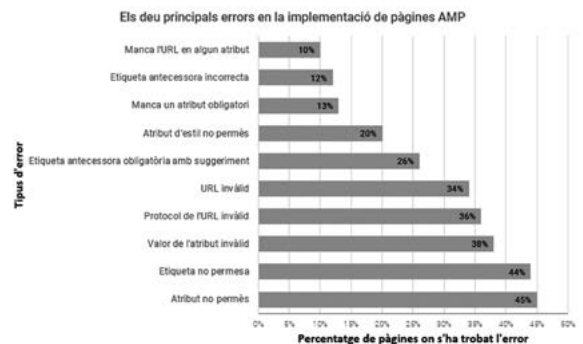


Figura 12. Percentatge d'aparició dels deu principals errors⁵⁷ en la implementació de pàgines AMP segons l'estudi de SEMrush.⁵⁸

52. <<https://validator.ampproject.org>>.

53. AMP4Email pretén portar la tecnologia AMP als correus electrònics, no només millorant-ne la velocitat, sinó també afegint-hi noves característiques interactives que permetran als usuaris interactuar amb tercers llocs web sense sortir del seu gestor de correu electrònic. Dues de les empreses que ja treballen en aquest nou estàndard són Pinterest i Doodle, els correus electrònics de les quals ens permetran afegir pins, escriure comentaris o participar en les enquestes que ens enviiïn els nostres contactes des del mateix missatge.

54. <<https://developers.google.com/web/tools/chrome-devtools/console/>>.

55. NPM és un gestor de paquets per treballar amb el llenguatge de programació JavaScript.

56. Elena Terenteva, «Top 10 AMP implementation mistakes: SEMrush Study» [en línia]. *SEMrush blog* (Oct. 2017), <<https://www.semrush.com/blog/amp-mistakes-semrush-study/>> [Consulta: 09/06/2018].

57. La llista completa d'errors es pot consultar a: <https://www.ampproject.org/es/docs/troubleshooting/validation_errors>.

58. Elena Terenteva, *op. cit.*

Des de la Google Search Console⁵⁹ també podem fer el seguiment dels errors detectats a les pàgines AMP durant el procés d'indexació de Google. Els errors s'organitzen entorn de deu categories diferents, que, al seu torn, es classifiquen en dos nivells de severitat: crítics, entre els quals s'hi inclouen tots aquells errors que impediran que la pàgina web sigui processada com a pàgina AMP vàlida; i no crítics, o advertències i recomanacions per millorar el codi font i fer-lo el més compatible i robust possible.

Issue	Severity ▲
1 Prohibited or invalid use of HTML Tag	Critical
2 AMP markup missing or incorrect	Critical
3 Prohibited HTML Tag with AMP equivalent	Critical
4 User-authored JavaScript found on page	Critical
5 Invalid CSS stylesheet	Critical
6 Major content mismatch	Critical
7 Use of deprecated tags or attributes	Non-critical
8 Minor content mismatch	Non-critical
9 Invalid structured data element	Non-critical
10 Add AMP structured data, if supported	Non-critical

Figura 13. Categories possibles d'errors amb els quals s'organitzen les incidències AMP a la Google Search Console.

L'origen de la majoria d'aquests errors el trobem en una incorrecta implementació de l'especificació AMP HTML per part dels sistemes de gestió de continguts que donen suport a la generació automatitzada d'aquestes pàgines a partir de l'HTML original. També, com passa habitualment, en el contingut generat pels usuaris editors dels llocs web sense coneixements tècnics.

2.6. Les AMP com a factor de posicionament

Tot i que les Google AMP no són per si mateixes un factor de posicionament, el seu ús sí que implica una reducció considerable en el temps de càrrega d'una pàgina web, un fet que pot millorar l'experiència d'usuari dels visitants i, per extensió, analítiques web com la taxa de rebot, la quantitat de pàgines vistes, o el temps de càrrega, aspectes que sí que influeixen en el rànquing de resultats.

D'altra banda, Google ha apostat clarament per donar preferència a les pàgines AMP entre els resultats de cerca oferts als usuaris que fan consultes des de mòbils. A més, entre aquests primers resultats de cerca, Google ha incorporat un nou bloc de resultats anomenat «històries destacades», que garanteix una gran visibilitat als resultats que es mostren d'aquesta manera.⁶⁰ Les notícies destacades poden aparèixer en format de carrusel horitzontal o en un bloc de tres notícies verticals, ambdues amb informació sobre el mitjà i amb una imatge destacada que ocupa una part preferent de la interfície.



Figura 14. El bloc de notícies destacades apareix tant a la versió per a ordinador de taula com a la versió mòbil de la pàgina de resultats de cerca de Google. En el cas de la versió mòbil, totes les notícies que apareixen al carrusel com a destacades disposen de pàgines AMP i tenen el marcatge Schema.org codificat en format JSON-LD. En els dos exemples mostrats, s'observa, davant del nom del mitjà i la data de publicació, el símbol de les AMP —un llamp— que n'identifica el tipus de resultat.

59. Google Search Console, conegut anteriorment com a Google Webmaster Tools, és un servei gratuït de Google que permet als propietaris d'un lloc web fer el seguiment del procés d'indexació del seu contingut i optimitzar-ne la visibilitat en el cercador de Google.
 60. Lin Pophal, «The state of mobile content» [en línia]. *eContent: the magazine of electronic research & resources*, v. 40, n. 1 (2017), p. 7-9, <<http://www.econtentmag.com/Articles/Editorial/Feature/The-State-of-Mobile-Content-2018-122570.htm>> [Consulta: 01/06/2018].

Per aspirar a ser candidat per aparèixer entre les històries destacades en la pàgina de resultats de cerca, no n'hi ha prou de disposar d'una pàgina AMP vàlida i un contingut rellevant per a la consulta de l'usuari, sinó que també cal incloure en la pàgina el marcatge Schema.org⁶¹ corresponent a les classes: Article, NewsArticle, BlogPosting i VideoObject codificat en format JSON-LD⁶² dins de la capçalera del document.

L'Schema.org és un vocabulari que permet definir les propietats amb les quals es poden codificar de manera semàntica els elements d'una pàgina web segons el tipus de continguts que presenti (notícies, apunts, receptes, esdeveniments, obres d'art, etc.).⁶³

```
<script type="application/ld+json">
{
  "@context": "http://schema.org",
  "@type": "NewsArticle",
  "mainEntityOfPage": "https://ampbyexample.com/samples_templates/",
  "headline": "Lorem Ipsum",
  "datePublished": "2016-04-21T11:55:02Z",
  "dateModified": "2016-04-21T11:55:02Z",
  "description": "A sample news article build with AMP.",
  "author": {
    "@type": "Person",
    "name": "Sebastian Benz"
  },
  "publisher": {
    "@type": "Organization",
    "name": "Google",
    "logo": {
      "@type": "ImageObject",
      "url": "http://cdn.ampproject.org/logo.jpg",
      "width": 600,
      "height": 60
    }
  },
  "image": {
    "@type": "ImageObject",
    "url": "https://ampbyexample.com/img/landscape_lake_1280x857.jpg",
    "height": 1280,
    "width": 857
  }
}
</script>
```

Figura 15. Exemple d'ús del marcatge Schema.org per codificar en format JSON-LD la classe NewsArticle.⁶⁴

En l'exemple anterior trobem propietats com el *Headline*, en què s'indica el titular de la notícia, la *image.url*, que serveix per enllaçar la imatge

destacada en el carrusel; el *publisher.logo*, que permet incloure el logotip del mitjà a sobre de la imatge destacada, i l'*author.name*, que indica el nom de l'autor de l'article, entre d'altres.

Més enllà d'aquest ús de l'Schema.org, Google recomana incorporar el marcatge propi d'aquesta especificació a la resta de la pàgina codificat en format de microdades.

3. Exemples d'implementació

Són precisament els mitjans de comunicació els primers que han adoptat l'especificació AMP per garantir-se una presència destacada en les pantalles dels usuaris mòbils. En són un bon exemple les principals capçaleres nacionals. Diaris com *Marca*, *El País*, *La Vanguardia* i *El Periódico* ja han adoptat aquest estàndard per garantir-se una posició privilegiada en els mòbils de l'audiència.

D'altra banda, la integració de les AMP en els models de negoci dels principals sectors tecnològics de la xarxa també s'està estenent cada vegada més. En aquest sentit, motors de cerca com Bing, el líder xinès Baidu i Yahoo Japan; plataformes socials com Twitter, LinkedIn, Pinterest, Tumblr i Reddit; els sistemes de gestió de continguts WordPress, Drupal i Canvas; els llocs web de comerç electrònic eBay, AliExpress i BigCommerce són només alguns exemples destacats d'actors que ja han incorporat d'una manera o altra aquesta tecnologia.

Com s'ha comentat al llarg de l'article, són fonamentalment els llocs web de contingut els que han adoptat primer l'estàndard AMP. En aquest sentit, a més dels llocs web de mitjans de comunicació, un important nombre de blogs

61. <<http://schema.org>>.

62. JavaScript Object Notation for Linked Data.

63. Andreu Sulé Duesa, «Schema.org, la millora de la visualització dels resultats en els cercadors i molt més [en línia]». *BID: textos universitaris de biblioteconomia i documentació*, n. 34 (2015). <<http://bid.ub.edu/34/sule.htm>> [Consulta: 28/05/2018].

64. <https://ampbyexample.com/samples_templates/news_article/>.

també han incorporat aquesta tecnologia. Pel que fa als blogs de biblioteques i arxius, són pocs els que l'han incorporada conscientment. Diem de manera conscient perquè la principal plataforma de blogs, WordPress.com, l'ha activat per defecte en els comptes de tots els seus clients.⁶⁵ Es tracta d'una implementació correcta, però molt senzilla, que simplifica la interfície fins al punt d'eliminar elements com ara el menú de navegació, tal com es pot veure al blog Education -Updates de la National Archives and Records Administration dels Estats Units d'Amèrica (figures 16 i 17).

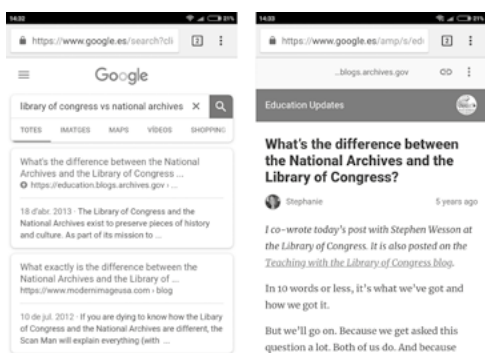


Figura 16. D'esquerra a dreta, pàgina de resultats de cerca de Google amb una entrada del blog Education Updates com a primer resultat i versió AMP d'aquesta mateixa entrada.

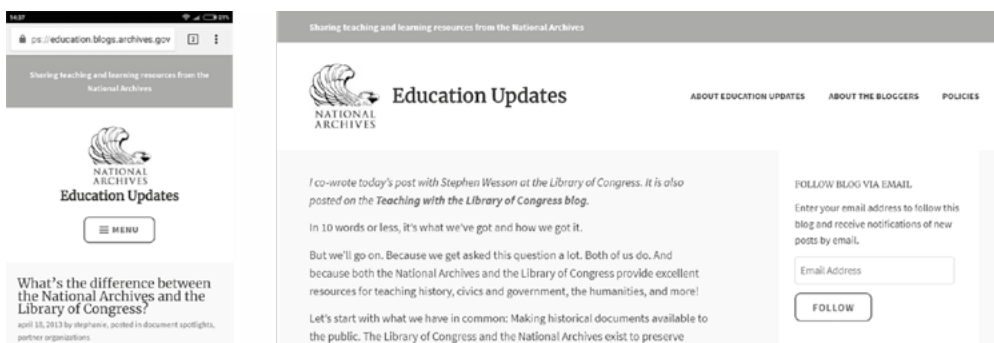


Figura 17. D'esquerra a dreta, es pot observar la mateixa entrada de la figura 16, però en la seva versió no-AMP vista des d'un mòbil i un ordinador de taula respectivament.

4. Conclusions

Al cap de tres anys d'aparèixer, les AMP han arribat a més de cinc bilions de pàgines web en més de trenta-un milions de dominis d'internet.⁶⁶ És un enorme grau d'adopció tenint present que encara es tracta d'una tecnologia jove i emergent. En aquest sentit, segons el cicle de sobreexpectació de la consultora tecnològica Gartner,⁶⁷ les Google AMP, una eina desenvolupada per representar la maduresa, l'adopció i l'aplicació social i consolidació d'una tecnologia, es troben actualment en el seu pic d'expectatives sobredimensionades i no assoliran l'altiplà de productivitat fins al període 2022-2027. Són, per tant, una tecnologia que encara ha de madurar molt per assolir l'objectiu de convertir-se en una autèntica alternativa per a les necessitats de tots els editors i consolidar-se com a estàndard per a la distribució de continguts en els dispositius mòbils, tenint en compte les propostes de Facebook i d'Apple, que també busquen dominar el sector de la distribució dels continguts a la xarxa.

65. «AMP (Accelerated Mobile Pages)». *WordPress.com. Support*. <<https://en.support.wordpress.com/amp-accelerated-mobile-pages/>> [Consulta: 17/08/2018].

66. Malte Ubl, «AMP's new horizons», *Accelerated mobile pages project* (2018), <<https://amphtml.wordpress.com/2018/02/13/amps-new-horizons/>> [Consulta: 24/06/2018].

67. Tuong Huy Nguyen, Megan Reinhart, *Hype cycle for mobile device technologies, 2017* [en línia]. Stamford, CT: Gartner, 2017, <<https://www.gartner.com/doc/3767664/hype-cycle-mobile-device-technologies>> [Consulta: 01/06/2018].

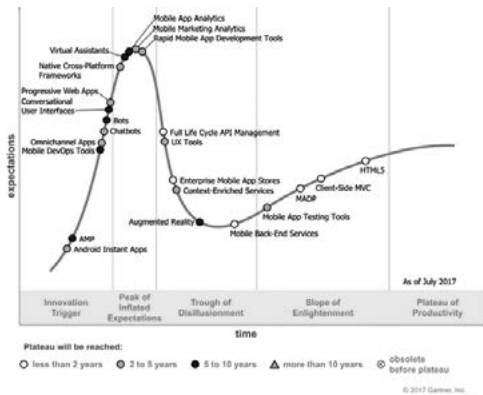


Figura 18. Cicle de sobreexpectació de les tecnologies associades als dispositius mòbils.⁶⁸

Les Google AMP són molt atractives en un context en el qual les visites des de dispositius mòbils ja han superat les que es produeixen des d'altres tipus de dispositius. Ho són especialment en el cas de petits i mitjans editors de continguts amb llocs web amb un rendiment molt pobre, i sense capacitat tècnica o econòmica per a corregir-ho. La facilitat d'implementació, sobretot mitjançant les principals plataformes de creació de continguts (WordPress, Drupal, etc.), la seva futura integració amb Gmail i altres mitjans socials i, sobretot, la necessitat de no perdre visibilitat en el principal punt d'accés dels continguts a internet, augura un escenari en el qual serà difícil obviar les AMP.

Per als editors que sí que tenen llocs web molt ben optimitzats per a la consulta des de mòbils, les Google AMP són molt menys atractives. Limitar l'ús del JavaScript i, amb això, les funcionalitats, el disseny i les característiques dels llocs web, suposa una important barrera

que impedeix que molts editors es decideixin per adoptar aquesta nova tecnologia.

Tanmateix, com s'ha comentat, el protagonisme que Google atorga al contingut AMP en la seva pàgina de resultats pot influir aquests editors en l'adopció de l'estàndard de Google.

Disposar d'un lloc web molt ben optimitzat no és l'única barrera per als editors de continguts. El fet de lliurar el contingut des d'un URL de Google és un dels principals motius de preocupació d'uns editors que, tot i tractar-se d'un projecte de codi obert, veuen com cada vegada tenen menys control dels seus continguts en un entorn en el qual les plataformes de distribució són les protagonistes. La incertesa de saber com aquesta tendència pot afectar en un futur els seus ingressos publicitaris és una altra qüestió que ha posat en alerta la indústria.

D'altra banda, el panorama de les tecnologies mòbils no es limita a les que hem descrit en aquest article. A les tendències que ja ens acompanyen de fa alguns anys, com el disseny web responsiu, les pàgines adaptatives, les versions mòbils i les més recents pàgines AMP, hem de sumar-hi altres opcions com les aplicacions web progressives⁶⁹ i les aplicacions mòbils instantànies,⁷⁰ que poden canviar dràsticament el panorama de les tecnologies mòbils els anys vinent, tot i que no són necessàriament excloents.⁷¹

Més enllà de les tecnologies que expliquem en aquest article, no hem d'oblidar les conseqüències derivades de l'ús dels dispositius mòbils com a principal dispositiu amb el qual

68. Tuong Huy Nguyen, Megan Reinhart, *op. cit.*

69. Les aplicacions web progressives (*progressive web apps*) són aplicacions desenvolupades amb estàndards web, però que es caracteritzen per presentar algunes de les funcionalitats, disseny i característiques pròpies de les aplicacions mòbils, com per exemple, la rapidesa, l'ús de notifikacions, la interfície adaptada a tasques específiques o que no faci falta disposar de connexió a internet per utilitzar-les. Tot això, sense la necessitat d'instal·lar-se als dispositius dels usuaris, sinó que es troben disponibles en els navegadors web.

70. Les aplicacions mòbils instantànies són una nova tecnologia que permet l'ús d'aplicacions mòbils sense necessitat d'instal·lar-les als dispositius.

71. Nancy Davis Kho, «Everything you need to know about progressive web apps» [en línia], *eContent: the magazine of electronic research & resources*, v. 41, n. 2 (2018), p 20-24, <<http://www.econtentmag.com/Articles/Editorial/Feature/Everything-You-Need-to-Know-About-Progressive-Web-Apps-124546.htm>> [Consulta: 02/06/2018].



A les tendències que ja ens acompanyen de fa alguns anys, com el disseny web responsiu, les pàgines adaptatives, les versions mòbils i les més recents pàgines AMP, hem de sumar-hi altres opcions com les aplicacions web progressives i les aplicacions mòbils instantànies, que poden canviar dràsticament el panorama de les tecnologies mòbils els anys vinents, tot i que no són necessàriament excel·lents.

es fan cerques a internet. En un futur pròxim, aspectes com les cerques geolocalitzades, amb un pes molt més gran del SEO local, o el creixement en l'ús de les cerques de veu, tindran un impacte molt gran en la manera com Google entendreà la rellevància dels continguts i en el disseny de la seva pàgina de resultats.

Agraïments i atribucions

Les icones utilitzades per crear algunes de les figures d'aquest article s'han obtingut del repositori d'icones The Noun Project⁷² i són obra d'AomAm, Kevin, StoneHub, Kamal i Wilson Joseph.

72. <<https://thenounproject.com>>.

Bibliografía

- ALCARAZ MARTÍNEZ, Rubén. «Webs i aplicacions mòbils: una aproximació des del punt de vista de l'experiència d'usuari» [en línia]. *Ítem: revista de biblioteconomia i documentació*, n. 61 (2016), p. 21-42. <<https://www.raco.cat/index.php/Item/article/view/317352>> [Consulta: 02/06/2018]
- BAKAUS, Paul. «Why AMP Cache exist» [en línia]. *Medium* (2017). <<https://medium.com/@pbakaus/why-amp-caches-exist-cd7938da2456>> [Consulta: 02/06/2018].
- DIGITAL in 2018: essential insights into Internet, social media, mobile, and ecommerce use around the world* [en línia]. [New York: Hootsuite, 2018]. <<https://hootsuite.com/uploads/images/stock/Digital-in-2018-001-Global-Overview-Report-v1.02-L.pdf>> [Consulta: 29/05/2018].
- ENCUESTA sobre equipamiento y uso de tecnologías de información y comunicación en los hogares* [en línia]. Madrid: Instituto Nacional de Estadística, 2017. <<http://www.ine.es/dynt3/inebase/es/index.htm?padre=3930>> [Consulta: 02/06/2018].
- ESTUDIO de Google Consumer Barometer 2017: el año de los móviles* [en línia] (Dic. 2017). <<https://www.thinkwithgoogle.com/intl/es-es/canales-de-publicidad/movil/estudio-de-google-consumer-barometer-2017-el-%C3%B1o-de-los-m%C3%B3viles/>> [Consulta: 31/05/2018].
- ERICSSON mobility report: on the pulse of the networked society* [en línia]. Stockholm: Ericsson AB, 2016. <<https://www.ericsson.com/assets/local/mobility-report/documents/2016/ericsson-mobility-report-feb-2016-interim.pdf>> [Consulta: 26/05/2018].
- FAR, Pierre. «Recommendations for building smartphone optimized websites» [en línia]. *Google webmaster central blog: official news on crawling and indexing sites for the Google index* (June 2012). <<https://webmasters.googleblog.com/2012/06/recommendations-for-building-smartphone.html>> [Consulta: 29/05/2018].
- FESSENDEN, Therese. «Five user requirements for online ads» [en línia]. *Nielsen Norman Group*. Articles (July 16, 2017). <<https://www.nngroup.com/articles/user-requirements-online-ads/>> [Consulta: 05/06/2018].
- FRICK, Tim. *Designing for sustainability: a guide to building greener digital products and services*. Sebastopol, CA: O'Reilly, 2016.
- GUALLAR, Javier. «Prensa digital 2015-2017: los medios frente a las plataformas tecnológicas» [en línia]. *Anuario ThinkEPI*, v. 12 (2018). <<https://recyt.fecyt.es/index.php/ThinkEPI/article/view/thinkepi.2018.33>> [Consulta: 29/05/2018].
- ILLYES, Gary. «What crawl budget means for Googlebot» [en línia]. *Google webmaster central blog: official news on crawling and indexing sites for the Google index* (Jan. 16, 2017). <<https://webmasters.googleblog.com/2017/01/what-crawl-budget-means-for-googlebot.html>> [Consulta: 25/05/2018].
- INFORME mobile en España y en el mundo 2017* [en línia]. Madrid: Ditrendia, 2017. <https://www.amic.media/media/files/file_352_1289.pdf> [Consulta: 2/06/2018].
- JIMÉNEZ IGLESIAS, Lucía; *et al.* «Experiencia de usuario y medios de comunicación: la regla de los tres clics en las webs de periódicos para smartphones» [en línia]. *Revista latina de comunicación social*, n. 73 (2018), p. 595-613. <<http://www.revistalatinacs.org/073paper/1271/30es.html>> [Consulta: 02/06/2018].
- KHO, Nancy Davis. «Everything you need to know about progressive web apps» [en línia]. *eContent: the magazine of electronic research & resources*, v. 41, n. 2 (2018), p. 20-24. <<http://www.econtentmag.com/Articles/Editorial/Feature/Everything-You-Need-to-Know-About-Progressive-Web-Apps-124546.htm>> [Consulta: 02/06/2018].
- KILLEEA, Patrick. *Web performance tuning*. Sebastopol, CA: O'Reilly, cop. 2002.
- KLOBOVES, Klemen. «Continuing to make the web more mobile friendly» [en línia]. *Google webmaster central blog: official news on crawling and indexing sites for the Google index* (March 16, 2016). <<https://webmasters.googleblog.com/2016/03/continuing-to-make-web-more-mobile.html>> [Consulta 27/05/2018].
- MAKINO, Takaki; Jung Chaesang; Phan, Doantam. «Finding more mobile-friendly search results» [en línia]. *Google webmaster central blog: official news on crawling and indexing sites for the Google index* (Feb. 26, 2015). <<https://webmasters.googleblog.com/2015/02/finding-more-mobile-friendly-search.html>> [Consulta: 27/05/2018].
- MEHTA, Ashish. «Engaging users through high quality AMP pages» [en línia]. *Google webmaster central blog: official news on crawling and indexing sites for the Google index* (Nov. 16, 2017). <<https://webmasters.googleblog.com/2017/11/engaging-users-through-high-quality-amp.html>> [Consulta: 29/05/2018].
- MEYERS, Peter J. «7 days after Mobilegeddon: how far did the sky fall?» [en línia]. *Moz blog* (Apr. 22, 2015). <<https://moz.com/blog/day-after-mobilegeddon>> [Consulta: 27/05/2018].
- «THE NEED for mobile speed: how mobile latency impacts publisher revenue» [en línia]. *Think with Google* (Sep. 2016). <<https://www.thinkwithgoogle.com/intl/en-154/insights-inspiration/research-data/need-mobile-speed-how-mobile-latency-impacts-publisher-revenue/>> [Consulta: 29/05/2018].
- NGUYEN, Tuong Huy; REINHART, Megan. *Hype cycle for mobile device technologies*, 2017 [en línia]. Stamford, CT: Gartner, 2017. <<https://www.gartner.com/doc/3767664/hype-cycle-mobile-device-technologies>> [Consulta: 01/06/2018].
- POPHAL, Lin. «The state of mobile content» [en línia]. *eContent: the magazine of electronic research & resources*, v. 40, n. 1 (2017), p. 7-9. <<http://www.econtentmag.com/Articles/Editorial/Feature/The-State-of-Mobile-Content-2018-122570.htm>> [Consulta: 01/06/2018].
- SINGHAL, Amit; CUTTS, Matt. «Using site speed in web search ranking» [en línia]. *Google webmaster central blog: official news on crawling and indexing sites for the Google index* (Apr. 9, 2010).

<<https://webmasters.googleblog.com/2010/04/using-site-speed-in-web-search-ranking.html>> [Consulta: 25/05/2018].

SOUDERS, Steve. *High performance web sites: essential knowledge for frontend engineers*. Sebastopol, CA [etc.]: O'Reilly, 2007.

— *Even faster websites*. Sebastopol, CA: O'Reilly, 2009.

SULÉ DUESA, Andreu. «Schema.org: la millora de la visualització dels resultats en els cercadors i molt més» [en línia]. *BID: textos universitaris de biblioteconomia i documentació*, n. 34 (2015). <<http://dx.doi.org/10.1344/BiD2015.34.23>> [Consulta: 28/05/2018].

LA TECNOLOGIA digital en España en 2018 [en línia]. [New York: Hootsuite, 2018]. <https://hootsuite-online-revenue.s3.amazonaws.com/Digital_in_2018_Local_country_report/DIGITAL_IN_2018_006_SPAIN_v1.01.pdf> [Consulta: 29/05/2018].

TERENTEVA, Elena. «Top 10 AMP implementation mistakes: SEMrush Study» [en línia]. *SEMrush blog* (Oct. 2017). <<https://www.semrush.com/blog/amp-mistakes-semrush-study/>>. [Consulta: 09/06/2018].

UBL, Malte. *AMP's New Horizons* [en línia]. Accelerated mobile pages project, 2018. <<https://amphtml.wordpress.com/2018/02/13/amps-new-horizons/>> [Consulta: 24/06/2018].

VARA MIGUEL, Alfonso; NEGREDO, Samuel; AMOEDO, Avelino. *Digital news report.es 2017: noticias en manos de la audiencia: las*

marcas periodísticas tradicionales referentes informativos para los internautas pese al auge de las redes sociales [en línia]. [Pamplona]: Universidad de Navarra, Facultad de Comunicación, Center for Internet Studies and Digital Life, 2017. <<https://drive.google.com/file/d/0B2eyawMqcpTyb2FDZUNUj0JmalE>> [Consulta: 30/05/2018].

WALTON, Philip. «User-centric performance metrics» [en línia]. *Web fundamentals: Google's opinionated reference for building amazing web experiences* (2018). <<https://developers.google.com/web/fundamentals/performance/user-centric-performance-metrics>> [Consulta: 26/05/2018].

WANG, Zhiheng; PHAN, Doantam. «Using page speed in mobile search ranking» [en línia]. *Google webmaster central blog: official news on crawling and indexing sites for the Google index*. (Jan. 17, 2018). <<https://webmasters.googleblog.com/2018/01/using-page-speed-in-mobile-search.html>> [Consulta: 25/05/2018].

WROBLEWSKI, Luke. *Mobile first*. New York, N.Y.: A Book Apart, 2011.

ZAKAS, Nicholas C. *High performance JavaScript: build faster web application interfaces*. Sebastopol, CA: O'Reilly; Yahoo! Press, 2010.

ZHANG, Fan. «Rolling out mobile-first indexing» [en línia]. *Google webmaster central blog: official news on crawling and indexing sites for the Google index* (March 26, 2018). <<https://webmasters.googleblog.com/2018/03/rolling-out-mobile-first-indexing.html>> [Consulta: 27/05/2018]. ■

Plans de Pensions de Caixa d'Enginyers

Atenció personalitzada per als membres del COL·LEGI OFICIAL DE BIBLIOTECARIS - DOCUMENTALISTES DE CATALUNYA

Volem crear amb tu el pla personal d'estalvi per a la teva jubilació.

T'ajudarem a planificar el teu estalvi a llarg termini per tal que aconseguis els teus objectius de jubilació.

Consulta les alertes de liquiditat i els indicadors de risc de cadascun dels nostres plans a <https://bit.ly/PlansPensionsCE>.

Si necessites més informació pots trucar a BancaTELFÒNICA al 902 300 321 (+34 93 310 26 26), de dilluns a divendres de 8:00 a 21:00h. Serà un plaer atendre't.

Tots els Plans de Pensions Individuals de Caixa d'Enginyers tenen com a Entitat Gestora: Caixa Enginyers Vida, SAU; com a entitat depositària i entitat promotora: Caixa d'Enginyers, SCC. Els Fons de Pensions estan auditats per Deloitte, SL -excepte el Caixa d'Enginyers 15, Fons de Pensions auditat per Auren Auditors SL-. Pots consultar el document de Dades Fonamentals per al Participi a <https://bit.ly/PlansPensionsCE> o a qualsevol oficina de Caixa d'Enginyers. Les rendibilitats passades no impliquen rendibilitats futures.

Grup  **Caixa d'Enginyers**
Som present. Som futur.



Som present. Som futur.



Marçal Rusiñol

Centre de Visió per Computador,
Departament de Ciències de la
Computació, Universitat
Autònoma de Barcelona
marcal@cvc.uab.cat

Article rebut l'octubre de 2018;
revisat el novembre de 2018.

Classificació semàntica i visual de documents digitals

Resum: En aquest article donem una visió de conjunt de la classificació automàtica de documents digitals. Veurem de quines maneres es poden descriure tant l'aparença visual com els continguts semàntics i textuais de documents, per poder entrenar models computacionals que siguin capaços de poder classificar, agrupar o fer cerques sobre documents digitals. Resumirem les tècniques clàssiques de la visió per computador i de processament del llenguatge natural, i veurem com els darrers avenços en l'aprenentatge profund (*deep learning*) han revolucionat aquest camp.

Paraules clau: Anàlisi de documents, visió per computador, processament del llenguatge natural, aprenentatge computacional, aprenentatge profund.

Clasificación semántica y visual de documentos digitales

Resumen: En este artículo presentamos una visión de conjunto sobre la clasificación automática de documentos digitales. Veremos de qué maneras se pueden describir tanto la apariencia visual como los contenidos textuales y semánticos de los documentos, para poder entrenar modelos computacionales que sean capaces de poder clasificar, agrupar o realizar búsquedas sobre documentos digitales. Resumiremos las técnicas clásicas de la visión por computador y de procesamiento del lenguaje natural, y veremos cómo los últimos avances en el aprendizaje profundo (*deep learning*) han revolucionado este campo.

Palabras clave: Análisis de documentos, visión por computador, procesamiento del lenguaje natural, aprendizaje computacional, aprendizaje profundo.

Semantic and visual classification of digitized documents

Abstract: This paper presents an overview of the problem of automatic classification of digitized documents. We will see the options available to describe both the visual appearance and the textual and semantic contents of these documents. We will review how these descriptions can be used for the classification, clustering or retrieval of digitized documents. We will summarise the state-of-the-art approaches both from the computer vision and natural language processing fields and will see how the latest breakthroughs in Deep Learning have revolutionized these fields.

Keywords: Document analysis, computer vision, natural language processing, machine learning, deep learning.



Introducció. La visió per computador i el processament del llenguatge natural

L'ús que fa de noves tecnologies la comunitat arxivística, bibliotecària i documentalista és una realitat en gairebé tots els passos de la cadena documental, des de les tècniques de conservació i els sistemes de digitalització fins a les eines informàtiques per a l'anotació de metadades o les bases de dades de consulta que faciliten la difusió de continguts. Tot i això, el procés documental requereix encara una gran quantitat de treball manual, cosa que pot arribar a convertir-se en un coll d'ampolla a l'hora de processar grans volums d'informació.

Gestionar aquests grans volums d'informació pot arribar a fer inviable, per motius de cost i temps, la tasca d'etiquetar i descriure acuradament cadascun dels documents amb les dades adequades. Com a conseqüència, gran part dels continguts dels fons documentals custodiats en biblioteques, hemeroteques o arxius estarà condemnada a romandre inaccessible als sistemes de cerca informàtica. En aquest sentit, un procés automàtic que pogués analitzar documents digitalitzats per descriure'n els continguts podria contribuir a facilitar-hi l'accés, a permetre'n la indexació automàtica i a fer accessibles els documents als motors de cerca.

La intel·ligència artificial ha esdevingut, en els darrers anys, una tecnologia emergent i ubíqua. L'abaratiment de la tecnologia, l'augment de la capacitat de càlcul dels ordinadors i l'accés a quantitats enormes de dades ho han fet possible. Avui dia, utilitzem serveis o productes basats en intel·ligència artificial de manera quotidiana. Quan fem una cerca als buscadors

d'internet; quan una web ens recomana articles que ens poden interessar basant-se en els nostres gustos; quan el telèfon mòbil ens avisa al matí que avui trobarem trànsit per arribar a la feina, o quan engeguem el nostre robot aspirador perquè recorri casa nostra mentre no hi som, es fan servir algorismes d'aprenentatge computacional.

En els darrers anys vivim una època daurada de la intel·ligència artificial, bàsicament per l'exploració dels mètodes d'aprenentatge profund¹ (en anglès, *deep learning*) que han revolucionat aquest àmbit, ja que obtenen rendiments fins i tot per sobre dels que assoleixen els humans. Els algorismes d'aprenentatge profund funcionen amb un sistema per capes, simulant el funcionament bàsic del cervell i les neurones. És a dir, el conjunt de capes que formen l'aprenentatge profund representen les neurones del cervell.

Dins del camp de la intel·ligència artificial i de l'aprenentatge profund prenen especial rellevància dues línies de recerca en particular. Per una banda, la visió per computador, que és la disciplina de la informàtica que fa que les màquines hi vegin i puguin analitzar i entendre els continguts d'imatges,² i, per l'altra, el processament del llenguatge natural, que és la disciplina informàtica que s'encarrega de tractar computacionalment el llenguatge humà.

En aquest article veurem quin és l'estat de l'art, tant emprant mètodes clàssics com analitzant els darrers mètodes punters en l'àmbit de l'aprenentatge profund, perquè es puguin dur a terme de manera automàtica els processos de classificació, agrupament i cerca de documents digitals. Començarem per descriure les tasques de classificació, agrupament i cerca per després donar una visió de conjunt dels mètodes que ens permeten descriure els documents di-

1. Yann LeCun, Yoshua Bengio, Geoffrey Hinton, «Deep learning» [en línia]. *Nature*, v. 521, n. 7553 (28 May 2015), p. 436-444 (28 May 2015). <<https://doi.org/10.1038/nature14539>> [Consulta: 20/09/2018].
2. Alicia Fornés; *et al.* «La visió per computador com a eina per a la interpretació automàtica de fons documentals». *Lligall: revista catalana d'arxivística*, n. 39 (2016), p. 18-44, <<https://www.raco.cat/index.php/lligall/article/view/340142/431080>> [Consulta: 20/09/2018].

Un procés automàtic que pogués analitzar documents digitalitzats per descriure'n els continguts podria contribuir a facilitar-hi l'accés, a permetre'n la indexació automàtica i a fer accessibles els documents als motors de cerca.

gitals, tant de manera visual, és a dir, quina és la seva aparença, com a partir dels seus continguts semàntics, és a dir, definint de què parlen.

1. Classificació, agrupament i cerca de documents digitals

Què entenem per classificació, agrupament i cerca d'elements? Quan parlem de classificació de documents³ entenem que prèviament s'ha definit un nombre finit de categories possibles i el que ha de fer l'algorisme és anar etiquetant de manera automàtica cada un dels elements a processar amb la categoria a la qual pertany. Es pot entendre, doncs, que el procés d'etiquetar els documents amb metadades a l'hora d'indexar-los per contingut es podria automatitzar amb un procés de classificació de documents. Per altra banda, els algorismes d'agrupació⁴ troben clústers de documents que siguin similars entre ells sense necessitat de definir prèviament el concepte de tipus de document. En aquest cas, només s'ha d'especificar, generalment, quin és el nombre de clústers diferents que vols que trobi el siste-

ma per agrupar els documents per similitud. Finalment, el procés de cerca de documents per similitud consisteix a involucrar l'usuari en el procés. Donada una col·lecció de documents digitals accessibles, qualsevol usuari pot venir a proposar una consulta. El sistema d'intel·ligència artificial haurà d'ordenar els documents de la base de dades de més a menys rellevants depenent de la consulta de l'usuari. En resum, en sistemes de classificació esperem que el sistema ens retorni etiquetes semàntiques donats els documents d'entrada; en sistemes d'agrupament, esperem que el sistema ens retorni els documents agrupats en diferents clústers significatius; i en sistemes de cerca, esperem que, donada una consulta, el sistema ens retorni una llista ordenada dels documents en funció de la rellevància. En qualsevol dels tres casos, una de les peces fonamentats és com fer-ho perquè la màquina tingui una representació semàntica dels continguts dels documents per a poder-los classificar, agrupar o ordenar de manera automàtica.

Cal tenir en compte també que, dins del camp de la intel·ligència artificial i de reconeixement de patrons, els diferents mètodes se separen en dues branques diferents.⁶ Per una banda, hi ha els mètodes supervisats i, per l'altra, els mètodes no supervisats. La diferència entre aquestes dues aproximacions és que en els mètodes supervisats la màquina aprèn a dur a terme la tasca observant anotacions fetes per humans que li han estat donades al començament de l'aprenentatge. Si, per exemple, volem programar un algorisme de classificació d'imatges que sigui capaç de diferenciar entre fotos de diferents animals, haurem de donar a la màquina un conjunt de fotos prèviament anotades i separades de manera manual perquè la màquina pugui aprendre quines són les caracte-

3. Simone Marinai, «Page similarity and classification», En: David Doermann, Karl Tomre (eds.). *Handbook of document image processing and recognition*, New York, [etc.]: Springer, 2014, p. 223-253.
4. Christopher D. Manning, Prabhakar Raghavan, Hinrich Schütze, «Cap. 16: Flat clustering». En: *Introduction to information retrieval*, Cambridge, [etc.]: Cambridge University Press, 2008, p. 349-375, Accés en línia: <<https://nlp.stanford.edu/IR-book/pdf/16flat.pdf>> [Consulta: 20/09/2018].
5. Ricardo Baeza Yates, Berthier Ribeiro Neto, *Modern information retrieval: the concepts and technology behind search*, Harlow, [etc.]: Addison-Wesley; Pearson, 2011.
6. Christopher Bishop, *Pattern recognition and machine learning*, New York, [etc.]: Springer, 2006.

rístiques visuals que ens diferencien entre fotos de gossos, gats, cavalls i zebres. Per altra banda, si el que volem és tenir un algorisme que ens agrupi una col·lecció d'imatges en funció de la seva similitud, no caldria proporcionar a la màquina tot d'etiquetes descrivint els diferents continguts de les imatges, sinó que la màquina haurà de trobar de manera autònoma i sense cap mena d'informació externa quina és l'agrupació que minimitza la distància entre imatges. En aquest cas, doncs, estariem parlant d'algorismes no supervisats.

Ara bé, tant si parlem d'algorismes de classificació, com d'agrupament o de cerca, la peça clau perquè els sistemes d'intel·ligència artificial es comportin adequadament és la manera com es representen els documents. Al cap i a la fi, les màquines només saben fer operacions numèriques i, per tant, hem de trobar la manera de poder representar els documents amb nombres. Aquestes representacions numèriques dels continguts dels documents, anomenades descriptors, han de poder permetre

que el càlcul de distàncies o similituds entre representacions tingui correlació amb la nostra percepció de la similitud dels documents. Així doncs, donat un document d'entrada, el descriptor extraurà característiques discriminatòries del document per a obtenir-ne una representació numèrica. Aquests descriptors hauran de seguir els principis següents per poder ser útils: si tenim dos documents semblants i en calculem els descriptors, la diferència (distància) entre aquests haurà de ser un valor petit, i contràriament, si calculem els descriptors de dos documents prou diferents, la seva distància haurà de ser elevada. Ara bé, aquesta noció de similitud entre documents pot ser molt variada depenent de l'aplicació final.

Per exemple, imaginem-nos que volem un sistema de classificació de documents en l'àmbit administratiu. Partim de la base que tenim una vintena de formularis i que, per tant, a l'hora de classificar els documents volem veure a quin tipus de formulari correspon el document digitalitzat. En aquest cas, com que les plantilles



estan definides prèviament i com que qualsevol instància d'un formulari concret tindrà la mateixa aparença visual, siguin quins siguin els continguts, necessitarem que la representació numèrica d'aquests documents tingui en compte l'aparença visual que tenen i no tant el contingut textual. D'altra banda, imaginem-nos que volem trobar agrupacions d'articles de diaris que parlin dels mateixos temes. En aquest cas, no ens interessarà quin format tenen aquests documents, però sí que ens interessarà definir la noció de similitud d'acord amb els continguts semàntics textuais dels documents que s'analitzen.

A continuació analitzarem aquestes dues modalitats de representació de documents digitals i veurem alguns exemples de sistemes automàtics de classificació i cerca de documents, tant per contingut com per aparença visual.

2. Descripció visual de documents

En aquest apartat ens centrarem a estudiar l'estat de la qüestió en les representacions numèriques de documents digitalitzats que codifiquin l'aparença visual dels documents.

2.1. Mètodes clàssics

Els mètodes basats en la visió per computador per a descriure l'aparença visual de documents digitalitzats processen els documents directament en format d'imatge. Sovint els descriptors més simples utilitzen estadístiques calculades a partir de funcions de baix nivell per codificar com es veuen els documents. Per exemple, Rusiñol, *et al.*⁷ proposa un descriptor de docu-

ments que codifica de manera jeràrquica, dividint les imatges de documents en una graella amb les densitats dels píxels en cada una de les cel·les. El vector numèric resultant descriu l'aparença visual del document, com si ens el miréssim de lluny.

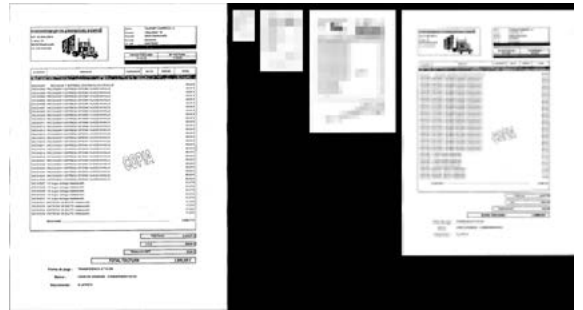


Figura 1. Exemple d'imatge de document i del descriptor jeràrquic visual proposat per Rusiñol, *et al.*

Aquests descriptors simples són útils quan es tracta de problemes en què els documents del mateix tipus són visualment similars encara que els continguts poden variar (p. ex., formularis).

Hi ha mètodes més elaborats que codifiquen la similitud del document en funció de la seva estructura. Les característiques estructurals s'obtenen a partir d'una anàlisi de disseny lògic o físic de les pàgines del document. L'anàlisi del disseny físic descompon les imatges del document en blocs i la similitud del document es pot expressar en termes de les relacions espacials entre aquests blocs. Com a exemple de família de mètodes que descriuen documents en termes de l'estructura dels elements lògics, citarem el treball presentat per Gordo.⁸ En aquest cas, la descripció del document codifica com es localitzen els elements lògics i quines són les relacions espacials que tenen.

7. Marçal Rusiñol, *et al.* «Multipage document retrieval by textual and visual representations». En: International Conference on Pattern Recognition (21st: 2012: Tsukuba, Japan). *Proceedings of the 21st International Conference on Pattern Recognition, November 11-15, 2012*. [Tsukuba, Japan: International Association for Pattern Recognition, 2012], p. 521-524, <<http://www.cvc.uab.es/~marcal/pdfs/ICPR12.pdf>> [Consulta: 20/09/2018].
8. Albert Gordo, *et al.*, «A kernel-based approach to document retrieval» [en línia]. En: IAPR International Workshop on Document Analysis Systems, Boston, Massachusetts, USA, June 09-11, 2010, *DAS'10: proceedings of the 9th IAPR International Workshop on Document Analysis Systems, Boston, Massachusetts, USA, June 09-11, 2010*, [New York: Association for Computing Machinery, 2010], p. 377-384, <<http://www.cvc.uab.es/~marcal/pdfs/DAS10b.pdf>> [Consulta: 20/09/2018].

Els descriptors estructurals són molt més eficients per avaluar la semblança visual entre els diversos tipus de documents que no els mètodes basats en imatges. No obstant això, tenen l'inconvenient que un mateix ha de computar la distància entre les dues estructures de disseny i això és computacionalment car.

2.2. Mètodes basats en aprenentatge profund

Els mètodes basats en aprenentatge profund per a la representació de l'aparença visual d'imatges de documents segueixen una estructura força similar. Un model computacional

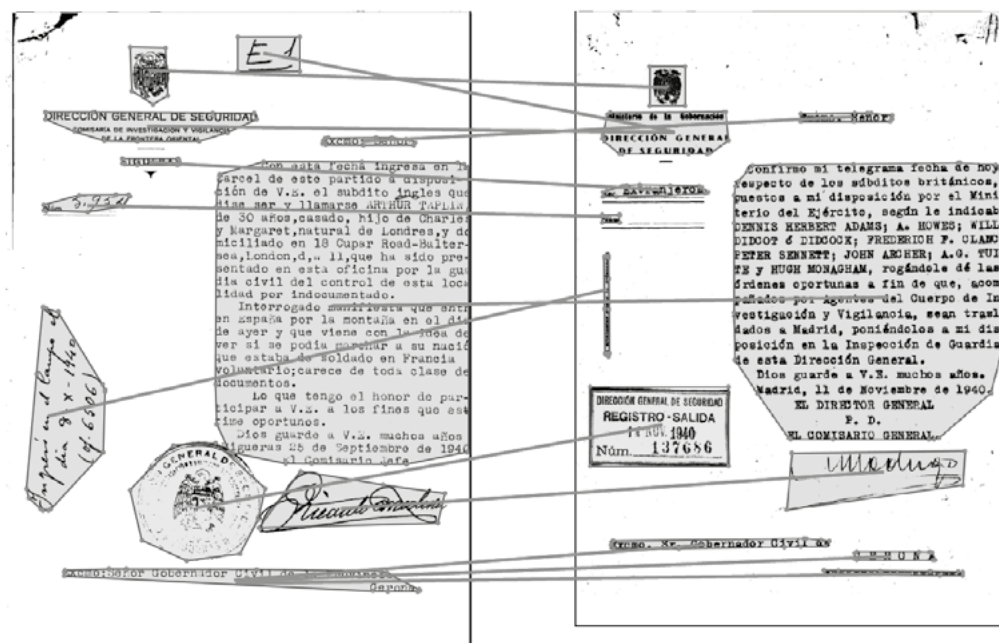


Figura 2. Exemple d'aparellament entre dos documents mitjançant descriptors d'estructura proposat per Gordo, *et al.*

De descriptors visuals de documents se n'han proposat centenars al llarg dels anys, cadascun amb els seus punts forts i punts febles. Recomanem al lector àvid de més detalls que faci un cop d'ull als articles de revisió de l'estat de la qüestió de Chen⁹ o de Doermann.¹⁰ Ara bé, com a la resta d'aplicacions de visió per computador, l'explosió de l'aprenentatge profund va revolucionar aquest àmbit i, avui dia, els descriptors d'imatges de documents basats en xarxes neuronals són els que predominen pel seu elevat rendiment.

conegut com a xarxa neuronal s'entrena amb un conjunt etiquetat de dades per a una tasca de classificació. És a dir, necessitem un conjunt força gran de documents etiquetats que es farà servir com a conjunt d'entrenament del sistema. Per poder entrenar de manera efectiva aquests models, normalment es necessiten centenars de milers o, fins i tot, milions d'imatges etiquetades. En aquest àmbit hi ha conjunts de bases de dades públiques i gratuïtes formades per documents prou heterogenis, cosa que permet que es puguin fer aquests entrenaments.

- Nawei Chen, Dorothea Blostein, «A survey of document image classification: problem statement, classifier architecture and performance evaluation» [en línia], *International Journal on Document Analysis and Recognition*, v. 10, n. (2006), p. 1-16, <<http://research.cs.queensu.ca/~blostein/IJDARSurvey2007.pdf>> [Consulta: 26/09/2018].
- David Doermann, «The indexing and retrieval of document images: a survey» [en línia], *Computer Vision Image Understanding*, v. 70, n. 3 (1998), p. 287-298, <<https://pdfs.semanticscholar.org/bbf9/b8ece2d8c392e91eded212af8174c7ab265.pdf>> [Consulta: 20/09/2018].

La xarxa neuronal, un cop entrenada perquè pugui classificar bé aquests documents, en realitat, ha calculat i extret un seguit de característiques que són les que permeten diferenciar un document d'un altre. Així doncs, independentment de quina tipologia d'imatges de documents es vulgui tractar, o de si es volen els descriptors tant per classificar com per agrupar o fer cerques, podem emprar aquestes xarxes neuronals ja entrenades amb el conjunt de dades públiques per calcular uns descriptors que són molt efectius i discriminatoris.

Per exemple, Csurka, *et al.*¹¹ van mostrar com es poden entrenar models de xarxes neuronals prou coneguts (com per exemple, AlexNet o GoogleNet) per poder representar característiques discriminatòries en imatges de documents. Aquests models, entrenats amb una base de dades de centenars de milers de documents, han après quines són les característiques visuals importants per poder-los discernir. El pas sorprenent és, però, que aquests models, un cop entrenats, es poden emprar per tractar lots de documents completament diferents dels mostrats en l'entrenament i, tot i així, aportar rendiments, tant en classificació com en cerca, molt superiors als prèviament proposats amb tècniques basades en anàlisi d'imatges.

3. Descripció semàntica dels continguts dels documents

En l'apartat anterior hem entrevist diferents maneres de poder descriure numèricament l'aparença visual dels documents. Ara bé, en molts casos ens interessarà descriure els documents en funció del tema que tracten, és a dir, dels continguts semàntics i textuals en lloc dels purament visuals. En aquest apartat, ens centrarem a estudiar l'estat de la qüestió en representacions numèriques de documents digitalitzats que codifiquin els continguts textuals. Començarem per parlar del reconeixement òptic de caràcters, en aquells casos en què els documents s'hagin hagut de digitalitzar amb un escàner i, per tant, d'entrada no els tinguem en format textual.

3.1. Reconeixement òptic de caràcters

Es pot considerar que l'origen de l'anàlisi de documents mitjançant la visió per computador apareix als anys seixanta, amb els primers sistemes de reconeixement òptic de caràcters (ROC). Els sistemes de ROC parteixen d'una imatge d'un document digitalitzat per produir un fitxer de text electrònic editable. Aquests

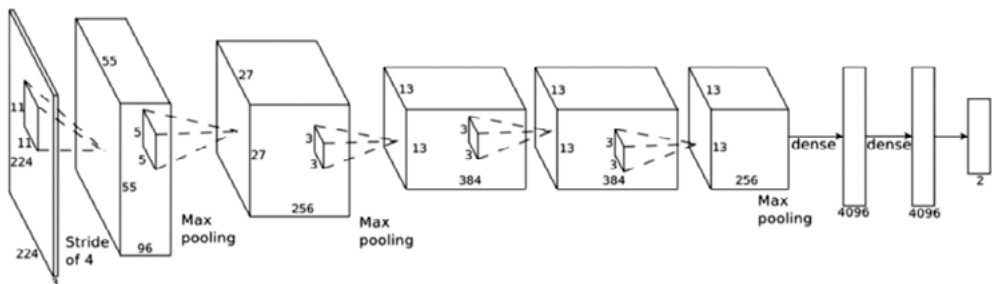


Figura 3. Exemple de l'arquitectura de la xarxa neuronal coneguda com a AlexNet, proposada per a codificar imatges de documents per Csurka, *et al.*

11. Gabriela Csurka, *et al.*, «What is the right way to represent document images?» [en línia], [arXiv preprint (Dec. 5, 2016)], <<https://arxiv.org/abs/1603.01076>> [Consulta: 20/09/2018].

sistemes integren, en primer lloc, un model de la forma de les lletres i un model lingüístic sobre les probabilitats que aquestes es combinin segons el llenguatge d'escriptura. Per tant, els programes de ROC reconeixen agrupacions de píxels com a lletres i, en un nivell superior, validen les interpretacions conjuntes per acabar transformant una imatge en un arxiu editable de paraules.

El programari de ROC ha evolucionat molt i avui dia té bones prestacions, especialment pel que fa a documents impresos i digitalització de qualitat. Les aplicacions d'ofimàtica i els escàners domèstics acostumen a incorporar un programari de ROC que permet transcriure automàticament els documents quotidians. Comercialment, grans corporacions com Nuance (OmniPage), Abbyy (FineReader) o Google (Tesseract) ofereixen bons sistemes que després altres empreses de serveis adapten a determinats escenaris, com ara el processament postal, la lectura de xecs bancaris i la incorporació de factures a sistemes de planificació de recursos empresarials (ERP).

Ara bé, tenir un fitxer amb el text electrònic automàticament extret, no acaba de solucionar l'accessibilitat sobre col·leccions de documents. Encara necessitem trobar com es poden codificar de manera numèrica continguts textuals per poder classificar documents semànticament. És a dir, poder tenir models computacionals que siguin capaços d'agrupar documents dependent dels conceptes semàntics que tractin.

3.2. De les estadístiques de paraules a la semàntica latent

Pel que fa a la descripció textual, el model de descripció més emprat es basa simplement en el càlcul d'estadístiques bàsiques sobre l'aparició de les diferents paraules en un text. Aquesta representació, coneguda com a «sacs de paraules» (*bag-of-words model*) ja s'emprava als anys cinquanta.¹² Tot i que són simples, aquests models funcionen prou bé en molts casos, encara que comencen a tenir limitacions a l'hora de tractar corpus de dades de certa envergadura. El principal problema radica en el fet que dos documents poden parlar de la mateixa cosa i fer servir paraules completament diferents.

Per fer front a aquesta limitació, apareixen tècniques computacionals que intenten extreure la semàntica latent que hi ha als documents. És a dir, els documents no es defineixen segons les paraules que hi apareixen, sinó segons els temes semàntics de què tracten i de quina manera els tracten. Aquestes tècniques, proposades a principis dels noranta per Deerwester, *et al.*,¹³ foren ràpidament adoptades i són a la base de tots els motors de cerca per internet. El seu punt fort radica en el fet que es poden recuperar documents encara que cap de les paraules formulades a la consulta apareguin al text original. Així es van solucionar els problemes que tenien els mètodes de sacs de paraules amb la sinonímia i la polisèmia. Aquests mètodes també tenen el gran avantatge de poder tractar textos de llargada arbitrària, amb la qual cosa es poden processar consultes sobre documents de diverses pàgines.¹⁴

12. Zellig Harris, «Distributional structure» [en línia], *Word*, v. 10, n. 2 (1954), p. 146-162, <<https://www.tandfonline.com/doi/pdf/10.1080/00437956.1954.11659520>> [Consulta: 20/09/2018].

13. Scott Deerwester, *et al.*, «Indexing by latent semantic analysis» [en línia], *Journal of the American Society for Information Science*, v. 41, n. 6 (1990), p. 391-407, <<http://lsa.colorado.edu/papers/JASIS.lsi.90.pdf>> [Consulta: 20/09/2018].

14. Albert Gordo, *et al.*, «Document classification and page stream segmentation for digital mailroom applications» [en línia], En: *International Conference on Document Analysis and Recognition (12th: 2013: Washington, DC), ICDAR 2013: proceedings of the 12th International Conference on Document Analysis and Recognition, Washington, DC, 25-28 August 2013*, Washington, DC: IEEE Computer Society, cop. 2013, <<http://www.cvc.uab.es/~marcal/pdfs/ICDAR13c.pdf>> [Consulta: 20/09/2018].



3.3. Mètodes basats en aprenentatge profund

Tal com ha passat en l'àmbit de la visió per computador, l'aparició de les xarxes neuronals també ha revolucionat el camp del processament del llenguatge natural. Tot i que les tècniques basades en l'anàlisi de la semàntica latent es continuen emprant pel seu bon rendiment i pel fet que no es necessiten dades etiquetades per entrenar-los, ja que es tracta de mètodes no supervisats, apareixen mètodes basats en l'aprenentatge profund que són molt prometedors, com és el cas del mètode de *word2vec* proposat per Mikolov, *et al.*¹⁵

El mètode *word2vec* és capaç de representar numèricament paraules capturant-ne el significat semàntic. El mètode es basa en una xarxa neuronal que, donades unes quantes paraules

L'aparició de les xarxes neuronals també ha revolucionat el camp del processament del llenguatge natural.

d'un text, ha de predir quina serà la paraula següent. Aquesta xarxa arriba a tenir una representació numèrica del significat de les paraules molt rica que, per tant, es pot emprar per a representar de què parlen els documents.

Com a exemple del poder discriminatori a escala semàntica del mètode *word2vec*, ens podem fixar en les relacions entre les representacions vectorials d'algunes paraules. Quan inspeccionem aquests casos, es fa evident que els vectors capturen informació semàntica

15. Tomas Mikolov, *et al.*, «Efficient estimation of word representations in vector space» [en línia]. En: International Conference on Learning Representations (2013: Scottsdale, AZ), *Proceedings of the International Conference on Learning Representations* (ICLR 2013), [s.l.: ICLR, 2013], <<https://arxiv.org/abs/1301.3781>> [Consulta: 20/09/2018].

general i, de fet, força útil de les paraules i les relacions que tenen entre elles. És interessant veure que determinades àrees en l'espai vectorial induït s'especialitzen cap a relacions semàntiques, com poden ser les relacions entre masculí i femení, de temps verbals o, fins i tot, relacions de país i capital entre les paraules, tal com s'il·lustra en la figura següent.

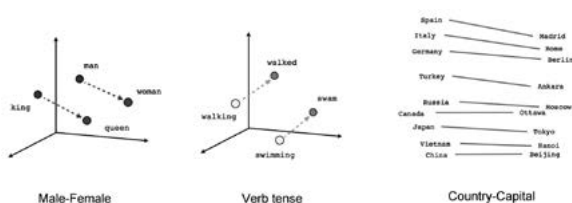


Figura 4. Exemple de relacions semàntiques entre paraules amb el mètode de *word2vec* de Mikolov, *et al.*

Com que amb el mètode de *word2vec* cada paraula es representa amb un nombre, es poden fer operacions aritmètiques entre paraules per poder visualitzar el poder de representació dels continguts semàntics, com per exemple:

Rei + (home - dona) = Reina

Què codifica una relació del tipus x és a y el que z és a...? Altres exemples d'aquestes relacions tenen a veure, tal com hem dit anteriorment, amb relacions país-capital, etc.

París + (França - Itàlia) = Roma

Einstein + (científic - pintor) = Picasso

Sushi + (Japó - Alemanya) = bratwurst

Microsoft + (Windows - Android) = Google

Aquestes representacions semàntiques numèriques tan riques en significat han revolucionat, de nou, el camp del processament del llenguatge natural, i no sols s'empren per a poder classificar o fer cerques semàntiques en fons documentals, sinó que també són en la base de sistemes de reconeixement de la parla o de traducció automàtica.

4. Conclusions

En els darrers anys s'ha avançat força tant en la visió per computador com en el processament del llenguatge natural, i en particular, en l'anàlisi de documents. L'avenç de les noves tecnologies, tant de programari com de maquinari, hi ha ajudat. Des de la recerca en l'àmbit de les enginyeries s'ha transferit coneixement que s'ha integrat a productes o serveis que avui dia estan integrats en arxius, biblioteques o plataformes de gran consum. Tanmateix, encara hi ha diversos reptes científics i tecnològics de cara a disposar de serveis de lectura universal. El que fa anys era el tractament de documents en estacions individuals, avui dia ha evolucionat cap a grans volums documentals (l'era de les dades massives). Els ciutadans del futur hauran de poder fer cerques per internet darrere de les quals hi haurà milions de documents provinents d'arxius d'arreu del món. Els algorismes han de ser prou eficients per no baixar el rendiment en aquesta nova dimensió. Lligada a aquest accés universal s'obre la necessitat de desenvolupament dels sistemes intel·ligents, és a dir, no només de reconeixement i transcripció literal, sinó també d'assoliment de la capacitat d'interpretació dels continguts. Només així els sistemes podran relacionar termes en diferents tipologies de documents i en diferents llengües. Aquest problema va més enllà de la visió per computador i requereix altres experteses d'intel·ligència artificial, com ara sistemes de modelatge i gestió del coneixement, representacions amb ontologies, etc.

Agraïments

Aquest treball ha estat finançat en part pel projecte TIN2014-52072-P, el programa CERCA de la Generalitat de Catalunya i el projecte aB-SINTHE de la Fundació BBVA. Agraïm a NVIDIA Corporation la donació de la GPU Titan Xp, que s'utilitza en el marc d'aquesta investigació.

**Plataforma de préstec en línia
d'audiovisuals**, cinema, sèries,
documentals, curts, concerts, animació
i cursos, creada exclusivament
per a les biblioteques.



efilm.online



Eloi FONT MANTÉ

Soci director, Font advocats
 eloifont@fontadvocats.com

Article rebut el novembre de 2018;
 revisat el desembre de 2018.

L'ús de les dades d'acord amb el Reglament General de Protecció de Dades

Resum: S'analitzen quins són els requisits jurídics que cal tenir en compte per utilitzar dades personals, especialment de manera massiva, d'acord amb el règim que preveu el nou Reglament General de Protecció de Dades (RGPD).

Paraules clau: Reglament General de Protecció de Dades, dades personals, dades massives (*big data*), règim jurídic, legitimitació.

El uso de los datos a la luz del Reglamento General de Protección de Datos

Resumen: Se analizan cuáles son los requisitos jurídicos a tener en cuenta para utilizar datos personales, especialmente de forma masiva, de conformidad con el régimen que prevé el nuevo Reglamento General de Protección de Datos (RGPD).

Palabras clave: Reglamento General de Protección de Datos, datos personales, datos masivos (*big data*), régimen jurídico, legitimitación.

The use of the data in accordance with the General Data Protection Regulation

Abstract: This article analyses the legal requirements that must consider in order to use personal data, especially in a massive way, in accordance with the regime established by the new General Data Protection Regulation (GDPR).

Keywords: General Data Protection Regulation (GDPR), personal data, big data, legal regime, legal basis.

... parser load (e, ...)
... r.e.style t; if (Ut.test(r)) return r; i s (v.supp
...); t.remove(); if (n == "none" || n == "") {Pt i.body, app
... type.html><html><body>"), Ht.close(); t Ht.body.appendC
... object"?t "")) """, i, n, r)}}; else if (n == v.type(t) == "obj
... gth; if (v.isFunction(n)) for (; u a; u++) r o u, s = \+ .tes
... l (c | u); f--> u a f (n, r, i), typeof u == "string" && (c
... n, n) n r | = t && ((s r ? e i | (i - {})) [r] = n [r]); i && v.exten
... , i = t && (i e.mimeType | n.getResponseHeader("content-ty
... (o), r[o]} function.On(e, t){var n, r, i, s, o = e.dataTypes.s
... + i | | a ["*." + i]; if (!n) for (r in a){s = r.split("."); if (s
... ", error: n?l: "No conversion from." + u + ". to." + i)}} u = i} re
... {} } function.\$n(){return.setTimeout(function(){qn = t},
... u = v.Deferred().always(function(){delete a.elem}), a = fu
... th(e, [f]), !1)}, f = u.promise({elem: e, props: v.extend({
... e, f.opts, t, n, f.opts.specialEasing [t] | | f.opts.easing)
... s; Qn(l, f.opts.specialEasing); for (; i < o; i++){r = Xn [i].ca
... ss(f.opts.progress).done(f.opts.done, f.opts.complete
... &(i = s [1], s = e [n] = s [0]), n !== r && (e [r] = s, delete e [n]), o =
... g = e.nodeType && Gt(e); n.queue || (l = v._queueHooks(e, "fx"
... .fire()}))}, e.nodeType === 1 && ("height" in t || "width" in
... lock": p.zoom = 1)), n.overflow && (p.overflow = "hidden", v.
... toggle"; if (s === (g? "hide": "show")) continue; m.push(r)}
... veData(e, "fxshow", !0); for (t in d) v.style(e, t, d[t]) }
... prototype.init(e, t, n, r, i)} function.Zn(e, t){var n, r = {h
... w: !1} var n, r, i = e.document, s = e.location, o = e.navigator
... ice, c = Array.prototype.indexOf, h = Object.prototype.toS
... 0) + | [\s\uFEFF\uxA0] + \$/g, w = /^(?: [^#<]* (<[\w\W]+) [^>]*
... /g, C = /^-ms-/, k = /-([\da-z])/gi, L = function(e, t){return
... tattachEvent("onreadystatechange", A), v.ready()}}, O = {};
... type of e == "string") {e.charAt(0) === "<" && e.charAt(e.le
... sPlainObject(n) && this.attr.call(e, n, !0), v.merge(this
... ctor(n).find(e)} return v.isFunction(e)? r.ready(e): (this
... this)}, get: function(e){return e == null? this.toArray(
... find"? r.selector = this.selector + (this.selector? " ": "
... (return e =+ e, e == -1? this.slice(e): this.slice(e, e + 1)
... , t.call(arguments).join(", ")}}, map: function(e){retu
... y, fn, v.extend = v.fn.extend = function(){var e, n, r, i, s,
... (e = arguments[0]) if (e == null || e == undefined) return []
... (e = Array.prototype.concat.apply([], arguments)) if (e == null || e == undefined) return []

Introducció

No descobrirem res de nou si comencem recordant al lector la disrupció tecnològica a què assistim i que alguns experts han batejat com una autèntica revolució, la «Revolució Industrial 4.0», en què les (noves) tecnologies esdevenen l'autèntic motor del desenvolupament de la societat.

La consultora Roca Salvatella i l'associació Adigital¹ expliquen molt bé com ha de ser el procés de transformació digital que les entitats i organitzacions han d'encarar per poder superar amb èxit els reptes de futur que se'ls plantegen. Aquest procés es divideix en quatre etapes: la primera, consistent en la mecanització i l'optimització dels processos, es va iniciar en sectors com la banca fa més de trenta anys, amb l'aparició i propagació dels caixers automàtics i les targetes de crèdit i dèbit o en el del taxi, amb l'ús de sistemes de navegació com el GPS. La segona etapa esdevé quan s'universalitza el punt de contacte amb el client, especialment amb l'aparició de les xarxes socials i les plataformes digitals de característiques similars, que obliguen les organitzacions a mecanitzar la gestió externa (*front-office*) per atendre els usuaris de manera permanent. La tercera se centra en el disseny de nous productes o serveis en què, per exemple, una empresa pot recomanar a un usuari, a través d'un bàner publicitari, la compra d'un producte concret, perquè uns dispositius automàtics (coneguts com a galetes o *cookies*) han detectat l'interès d'aquell usuari pel producte, en haver-lo consultat o cercat al seu propi web o en motors de cerca amb caràcter previ. I, finalment, la quarta etapa, que és aquella en què es reconsideren de dalt a baix els models de negoci establerts, a través de propostes plantejades pel que es coneix com a economia col·laborativa (Spotify

en el sector de la música, Airbnb en el del turisme o Uber en el transport en serien alguns exemples coneguts).

En aquest context de tecnificació i digitalització massiva, les dades han esdevingut la baula central de la cadena i és per aquest motiu que, sovint, se les defineix com «el petroli del segle XXI», atès que les organitzacions basen els seus models de negoci en la captació, anàlisi i explotació de les dades dels seus clients o usuaris. És per això que la regulació i el tractament de les dades, i especialment la de les dades de caràcter personal, esdevé cabdal ja que el dret a la privacitat dels usuaris pot veure's afectat.

A continuació ens proposem posar en relleu quins són els principals requisits que cal tenir en compte per poder fer ús de les dades, especialment a escala massiva, d'acord amb el nou Reglament General de Protecció de Dades, aplicable d'ençà del 25 de maig del 2018 i que suposa, per primera vegada, que tots els estats membres de la Unió Europea tenen idèntica regulació sobre aquesta matèria.

1. Marc jurídic del Reglament General de Protecció de Dades

El Reglament General de Protecció de Dades² (RGPD) suposa una modernització més que necessària de la normativa de protecció de dades vigent fins ara amb l'objectiu de permetre als ciutadans un millor control de les seves dades. També ha de permetre a les empreses aprofitar les oportunitats que es deriven del mercat únic digital.

1. Pepe Cerezo, Carlos Magro, Josep Salvatella, *Sobre la transformació digital i su impacto socioeconómico: 10 claves para el debate* [en línia], Madrid: RocaSalvatella; adigital, 2014, <<https://www.adigital.org/media/claves-tranformacion-digital.pdf>> [Consulta: 02/10/2018].
2. Unió Europea, *Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo, de 27 de abril de 2016, relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos y por el que se deroga la Directiva 95/46/CE (Reglamento general de protección de datos)*. En: *Diario Oficial de la Unión Europea*, L 119/1 (04/05/2016), p. 1-88, <<https://eur-lex.europa.eu/eli/reg/2016/679/oj>> [Consulta: 02/10/2018].

L'aparició de les xarxes neuronals també ha revolucionat el camp del processament del llenguatge natural.

El RGPD deroga la Directiva 95/46/CE³ del Parlament Europeu i del Consell, de 24 d'octubre de 1995, relativa a la protecció de les persones físiques pel que fa al tractament de dades personals i a la lliure circulació d'aquestes dades. La principal diferència entre els dos textos rau en la seva naturalesa jurídica. El RGPD és obligatori en tots els seus elements i directament aplicable en tots els estats membres de la Unió

Europea (UE). L'elecció d'aquest instrument jurídic no és intranscendent i respon a la necessitat que hi hagi més harmonització europea. La Directiva 95/46/CE, en canvi, va requerir, en el seu dia, la corresponent transposició que, a l'Estat espanyol, es va dur a terme mitjançant la Llei Orgànica 15/1999, de 21 de desembre, de protecció de dades de caràcter personal (LOPD).⁴ No obstant això, cal tenir en compte que el RGPD no va derogar de facto l'antiga LOPD, ni tampoc deroga la reforma d'aquesta, que ha estat aprovada recentment, el 5 de desembre de 2018,⁵ i s'adapta al règim del RGPD i que és aplicable en tot allò que no regula el propi reglament.

Pel que fa a l'àmbit territorial, el més rellevant és que el RGPD s'aplica al tractament de da-

3. Unió Europea, Directiva 95/46/CE del Parlamento Europeo y del Consejo, de 24 de octubre de 1995, relativa a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos. En: Diario Oficial de la Unión Europea, L 281 (23/11/1995), p. 31-50, <<https://eur-lex.europa.eu/eli/dir/1995/46/oj>> [Consulta: 02/10/2018].
4. Espanya, *Ley Orgánica 15/1999, de 13 de diciembre, de protección de datos de carácter personal*. En: *Boletín Oficial del Estado*, n. 298 (14/12/1999), p. 43.088-43.099, <<https://www.boe.es/eli/es/lo/1999/12/13/15>> [Consulta: 02/10/2018].
5. Espanya, *Ley Orgánica 3/2018, de 5 de diciembre, de protección de datos personales y garantía de los derechos digitales*. En: *Boletín Oficial del Estado*, n. 294 (6/12/2018), p. 119.788-119.857, <<https://www.boe.es/eli/es/lo/2018/12/05/3>> [Consulta: 02/10/2018].

des personals d'interessats que resideixin a la UE fet per organitzacions no establertes a la UE sempre que les activitats estiguin relacionades amb:

- L'oferta de béns o serveis a interessats residents a la UE, independentment de si se'ls requereix que els paguin.
- El control del seu comportament, en la mesura que aquest tingui lloc a la UE.

2. Les dades de caràcter personal i les dades massives

D'acord amb la normativa vigent en cada moment, el concepte de *dada personal* sempre ha tingut un doble vessant: ha de versar sobre persones físiques i aquestes han de ser identificades o, fins i tot, identificables.

A l'efecte, la Directiva 95/46/CE ja considerava que podien ser identificables les persones la identitat de les quals podia determinar-se, de manera directa o indirecta, «mitjançant números d'identificació o un o diversos elements específics característics de la identitat física, fisiològica, psíquica, econòmica, cultural o social».

El RGPD manté la mateixa definició, per bé que tracta d'aclarir-la amb alguns exemples del que es pot considerar «identificador» (nom, número d'identificació, dades de localització o identificador en línia), i es consolida així un concepte que cal entendre en sentit ampli.

Si bé és cert que ja fa temps que les organitzacions prenen bona part de les decisions basant-se en l'anàlisi de les dades, el que realment és nou és el volum de dades que es poden arribar a manegar i els canals o vies de captació. En bona mesura per l'aparició d'in-

ternet i, especialment, gràcies a l'eclosió de les xarxes i les plataformes digitals, que aboquen milions de gigabytes cada dia. Això no només es mantindrà en el temps, sinó que creixerà exponencialment amb tecnologies com la cadena de blocs (*blockchain*) o l'internet de les coses (*internet of things*), entre d'altres.

En aquest context, cal tenir en compte les activitats de dades massives (*big data*), que no són altra cosa que la tècnica que permet tractar i analitzar grans volums de dades, estructurades i no estructurades, de diferents formats i que, com apunta Elena Gil,⁶ moltes vegades provenen de fonts diferents, amb l'objectiu de poder atorgar-los una utilitat que proporcioni valor com, per exemple, descobrir patrons de comportament dels clients d'una organització per a crear publicitat dirigida més efectiva, predir tendències econòmiques o descobrir relacions anteriorment desconegudes entre variables que poden obrir les portes a la innovació. Cal tenir en compte que les dades massives poden col·lidir amb el principi de minimització de dades que expliquem en l'apartat 5 de l'article.

3. El consentiment com a principi general

El RGPD introdueix dues novetats essencials pel que fa al consentiment com a principi del tractament o processament de les dades personals:

- La necessitat d'una declaració o una clara acció afirmativa de l'interessat.
- La transformació del principi general del consentiment amb excepcions en una simple base de licitud per al processament de dades.

El consentiment de l'interessat ha de ser interpretat com la manifestació de voluntat lliure,

6. Elena Gil González, *Big data, privacidad y protección de datos* [en línia], Madrid: Agencia Española de Protección de Datos; Agencia Estatal Boletín Oficial del Estado, 2016, <<https://www.aepd.es/media/premios/big-data.pdf>> [Consulta: 02/10/2018].



11:10

Hem

Join us in Parliament Square from 11am for the historic unveiling of suffragist leader Millicent Fawcett's statue - the first of a woman in this iconic location. Bring your friends...

amanda abbington Retweeted



Marshall Mathers
How come your shirt is so bloody?

#FRAMED VIDEO OUT NOW - already out FramedOfficial...



0:17 6057 views

específica, informada i inequívoca, per la qual l'interessat accepta, sigui mitjançant una declaració o una clara acció afirmativa, el tractament de dades personals que el concerneixen.

El primer element a destacar és la necessitat d'una declaració o una clara acció afirmativa, de manera que ja no és possible acceptar consentiments tàcits o implícits derivats de la inactivitat de l'interessat com a base de licitud. O dit amb altres paraules, el consentiment inequívoc és aquell que s'ha prestat mitjançant una manifestació de l'interessat o mitjançant una clara acció afirmativa i, a diferència de la LOPD, no s'admeten consentiments tàcits o per omissió, basats en la inacció. En determinades situacions, el RGPD exigeix que el consentiment, a més d'inequívoc, sigui explícit, com succeeix en el cas del tractament de dades sensibles, en l'adopció de decisions automatitzades i en les transferències internacionals.

Els tractaments iniciats abans de l'aplicació del RGPD sobre la base del consentiment continuaran sent legítims sempre que aquest consentiment s'hagués prestat de la manera que preveu el mateix RGPD, és a dir, mitjançant una manifestació o acció afirmativa. En cas contrari, caldrà demanar a l'interessat un consentiment inequívoc o bé fonamentat el tractament i processament de dades en una altra base lícita que ho justifiqui.

El segon canvi és formal i interpretatiu, ja que el redactat que el RGPD atorga al consentiment ja no el situa com un principi stricto sensu de la protecció de dades, sinó com una mera base de licitud, que legitima o fonamenta el tractament o processament de dades. Cal recordar que la LOPD ens deia que «el tractament de les dades de caràcter personal requereix el consentiment inequívoc de l'afectat, llevat que la llei disposi una altra cosa». És a dir, una exigència general d'obtenir el consentiment, llevat d'excepció legal.

4. Altres vies de legitimació dels tractaments

El RGPD preveu un seguit de bases lícites per tractar les dades personals a part del consentiment. Més concretament, apunta la relació contractual, els interessos vitals de l'interessat i d'altres persones, l'existència d'una obligació legal per al responsable, l'interès públic o exercici de poders públics i els interessos legítims que prevalen del responsable o de tercers als quals es comuniquen les dades.

4.1 La relació contractual

El RGPD estableix que, quan el tractament és necessari per a executar un contracte en què l'interessat n'és part (o per a aplicar a petició d'aquest mesures precontractuals), la recollida de dades personals s'ha de comunicar a l'interessat fent constar una referència al contracte, específica i clara, que en cap cas pugui generar confusió. Es tracta d'una base de legitimació que, fins ara, es recollia en el marc legislatiu espanyol com una excepció al consentiment en cas d'existir una relació de negocis.

4.2 Interessos vitals de l'interessat o d'altres persones

El RGPD també preveu que és lícit el tractament en situacions especials, urgents o sobrevingudes, tot i que no constitueix una causa de legitimació general dels tractaments. La LOPD ja preveia una excepció al consentiment quan el tractament de les dades tenia com a finalitat protegir l'interès vital de l'interessat i enumerava dos supòsits: quan el tractament és necessari per prevenir o fer un diagnòstic mèdic o prestar assistència sanitària; quan l'afectat estigui físicament o jurídicament incapacitat per donar el seu consentiment.

4.3 Obligació legal per al responsable

El RGPD determina que també serà lícit el tractament quan hi hagi una obligació legal per al responsable, és a dir, un mandat que deriva del dret de la UE o de l'ordenament jurídic intern. En aquest cas, quan es demanen les dades personals s'haurà de fer constar de manera clara i simple quina és la norma que imposa aquesta obligació de processament i tractament de dades i informar-ne l'interessat.

4.4 Interès públic o exercici de poder públics

El text comunitari també inclou com a base de legitimació del tractament de dades el compliment d'una missió duta a terme en interès públic o en l'exercici dels poders públics conferits al responsable del tractament. A més, segons l'article 43 del RGPD, aquest cas pot ser interpretat en el sentit de limitar l'ús del consentiment en aquest àmbit, en la mesura que determinats tractaments fets per les autoritats públiques no admeten una posició lliure des de l'òptica del ciutadà o administrat, que es troba en una situació de desequilibri enfront del poder públic. En vista d'aquest raonament, el Grup d'Experts de la Unió Europea (GT29) va considerar en el seu moment que hi havia altres bases jurídiques que justificaven que les autoritats públiques poguessin tractar dades personals, i una d'aquestes bases és la necessitat de complir amb una missió en interès públic o en exercici de poders públics que li han estat conferits al responsable del tractament, quan derivi d'una competència atribuïda per una norma.

4.5 Interès legítim

Es tracta de la darrera base lícita per al tractament de les dades personals. No té marge ni recorregut quan el responsable del tractament és una administració pública, ja que aquestes ja tenen la base de legitimació de l'interès públic. No obstant això, l'interès legítim es planteja

ja com una base legítima per al tractament de dades que duen a terme empreses i organitzacions privades quan els interessos legítims perseguits pel responsable del tractament o per un tercer prevalen sobre els drets i les llibertats fonamentals de l'interessat.

L'interès legítim es contextualitza sempre en el marc de la relació entre el titular de les dades i el responsable del tractament. El RGPD determina que s'ha d'avaluar i justificar amb detall l'interès legítim, fins i tot en aquells supòsits més evidents. És fonamental fer una ponderació i un breu judici de proporcionalitat (idoneïtat i necessitat), en la qual es posa en la balança l'interès legítim del responsable o la raó que aquest esgrimeix per dur a terme el tractament, en una banda, i els drets i llibertats fonamentals de l'interessat, en l'altra.

En aquesta mateixa línia, es pot consultar el Dictamen 06/2014 del GT29, en què es tracta el concepte d'interès legítim del responsable del tractament, i que assenyala que sempre s'ha de ponderar el suposat *interès legítim del responsable del tractament* o tercers i els interessos o drets fonamentals de la persona interessada per determinar, en primer lloc, si existeix un interès legítim, i en segon lloc, si preval sobre els drets i interessos del titular de les dades. El GT29 enumera tot un seguit de factors rellevants a tenir en compte en el judici de proporcionalitat o ponderació: la naturalesa i l'origen de l'interès legítim, així com la necessitat i l'interès públic o de la comunitat afectada; la repercussió per a l'interessat i les seves expectatives raonables sobre què passarà amb les seves dades, la seva naturalesa i la forma de tractament, i les garanties que poden limitar un ús indegut (minimització de les dades, les tecnologies de protecció de la intimitat, l'augment de la transparència, el dret general d'exclusió voluntària i la portabilitat de les dades).

El mateix Dictamen 06/2014 va proposar al legislador europeu que el RGPD exigís que el responsable del tractament documenti la seva valoració d'acord amb els factors assenyalats

abans de procedir al tractament, amb l'objectiu de tenir més responsabilitat proactiva sobretot en els casos més complexos. El legislador europeu no va considerar aquesta proposta. No obstant això, l'AEPD (en les guies que ha publicat sobre l'aplicació del RGPD) sí que ha explicat que quan es recullen i es tractin dades d'acord amb un interès legítim, s'haurà de justificar sempre aquest interès o raó del tractament de les dades.

5. El principi de minimització de dades

Segons aquest principi, la recollida de dades personals ha de ser proporcional i, per aquest motiu, les dades han de ser necessàries, pertinents, adequades o idònies, i el seu tractament ha d'estar justificat.

Per tant, la recollida i el tractament de les dades personals ha de ser necessària per a la finalitat que es persegueix, no ha d'existir cap altra via per assolir la finalitat que sigui menys lesiva i, a més, ha de ser idònia, és a dir, capaç d'aconseguir el propòsit perseguit.

6. Decisions automatitzades

El RGPD també recull el dret de no ser objecte d'una decisió basada únicament en el tractament automatitzat, incloent-hi l'elaboració de perfils, que produeixi efectes jurídics en l'interessat o l'afecti significativament de manera similar.

S'entén per «elaboració de perfils» l'avaluació d'aspectes personals per a elaborar prediccions sobre una persona, fins i tot si no es prenen decisions. Únicament es prenen decisions automatitzades quan s'usen mitjans tecnològics per a decidir sense la intervenció de cap persona. L'elaboració de perfils i les decisions automatitzades són habituals en els sectors financers i bancaris, i cada vegada més s'estenen a altres àmbits i sectors.

Els responsables del tractament han de tenir en compte el dret a la protecció de dades quan es desenvolupin i dissenyin els productes.

Les decisions automatitzades poden estar permeses excepcionalment si una llei concreta preveu les garanties adequades o quan la decisió és necessària per formalitzar o executar un contracte, sempre que no hi hagi una altra manera d'aconseguir el mateix objectiu, o si l'interessat ha donat el consentiment explícit. En tots aquests casos, el responsable ha d'adoptar les mesures adequades per salvaguardar els drets i les llibertats i els interessos legítims de l'interessat, i haurà de garantir, com a mínim, el dret a obtenir intervenció humana del responsable, a expressar el seu punt de vista i a impugnar la decisió automatitzada.

Les decisions basades en algorismes o automatitzades no poden utilitzar categories especials de dades, llevat que l'interessat hagi donat el consentiment explícit o que ho permetin la legislació de la UE o la legislació nacional.

7. *Privacy by design*

Els sistemes de privadesa des del disseny són aquells en què les tecnologies o els dispositius s'ideen i desenvolupen tenint en compte la necessitat de la protecció de la privacitat; mentre que els sistemes de privacitat per defecte garanteixen que la configuració del dispositiu establerta de fàbrica sigui el més protectora i segura possible.

El RGPD incorpora aquests principis, que suposen una de les grans novetats en la normativa; de fet, l'article 78 del RGPD estableix que la protecció dels drets i les llibertats en el tractament de dades personals exigeix l'adopció de mesures tècniques i organitzatives apropiades per a garantir el compliment de la llei, sempre

tenint en compte l'estat de la tècnica. En concret, els responsables del tractament han de tenir en compte el dret a la protecció de dades quan es desenvolupin i dissenyin els productes. Com a possibles mesures a aplicar, el RGPD assenyala, a manera d'exemple, la minimització de dades, la reducció al màxim de les dades personals que se sol·liciten per a la prestació del servei, la pseudonimització de les dades personals quan la identificació no sigui necessària o raonable i la transparència en les funcions i en el tractament de les dades.

8. Avaluacions d'impacte

Per acabar, interessa destacar que el RGPD preveu l'obligació de dur a terme una auditoria específica, denominada *avaluació d'impacte*, quan el tractament de dades implica un alt risc per als drets i les llibertats de les persones, i això pot ser necessari en casos com quan es fan elaboracions de perfils sistemàtiques i exhaustives o tractaments massius de dades sensibles.

Aquesta auditoria ha d'incloure, com a mínim, una descripció de les operacions de tractament, la necessitat i proporcionalitat dels tractaments, una avaluació dels riscos i les mesures previstes per afrontar-los, incloent-hi les garanties, les mesures de seguretat i els mecanismes que garanteixen la privadesa.

Bibliografia

CEREZO, Pepe; MAGRO, Carlos; SALVATELLA, Josep. *Sobre la transformación digital y su impacto socioeconómico: 10 claves para el debate* [en línia]. Madrid: RocaSalvatella; adigital, 2014. <<https://www.adigital.org/media/claves-tranformacion-digital.pdf>> [Consulta: 02/10/2018].

ESPAÑA. *Ley Orgánica 15/1999, de 13 de diciembre, de protección de datos de carácter personal*. En: *Boletín Oficial del Estado*, n. 298 (14/12/1999), p. 43.088-43.099. <<https://www.boe.es/eli/es/lo/1999/12/13/15>> [Consulta: 02/10/2018].

— *Ley Orgánica 3/2018, de 5 de diciembre, de protección de datos personales y garantía de los derechos digitales*. En: *Boletín Oficial del Estado*, n. 294 (6/12/2018), p. 119.788-119.857. <<https://www.boe.es/eli/es/lo/2018/12/05/3>> [Consulta: 02/10/2018].

GIL GONZÁLEZ, Elena. *Big data, privacidad y protección de datos* [en línia]. Madrid: Agencia Española de Protección de Datos; Agencia Estatal Boletín Oficial del Estado, 2016. <<https://www.aepd.es/media/premios/big-data.pdf>> [Consulta: 02/10/2018].

UNIÓ EUROPEA. *Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo, de 27 de abril de 2016, relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos y por el que se deroga la Directiva 95/46/CE (Reglamento general de protección de datos)*. En: *Diario Oficial de la Unión Europea*, L 119/1 (04/05/2016), p. 1-88. <<https://eur-lex.europa.eu/eli/reg/2016/679/oj>> [Consulta: 02/10/2018].

— *Directiva 95/46/CE del Parlamento Europeo y del Consejo, de 24 de octubre de 1995, relativa a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos*. En: *Diario Oficial de la Unión Europea*, L 281 (23/11/1995), p. 31-50. <<https://eur-lex.europa.eu/eli/dir/1995/46/oj>> [Consulta: 02/10/2018]. ■



Pedro RUEDA RAMÍREZ

Departament de Biblioteconomia,
Documentació i Comunicació
Audiovisual, Universitat de
Barcelona
pedrorueda@ub.edu

Dolors SAUMELL CALAF

Biblioteca Pública de Tarragona
dsaumell@gencat.cat

Article rebut l'octubre de 2018;
revisat el novembre de 2018.

Les marques de propietat dels llibres de la Biblioteca d'Escornalbou: un patrimoni recuperat

Resum: La recuperació del patrimoni bibliogràfic d'Escornalbou s'ha fet basant-se en exlibris i *olim*; s'han identificat 570 registres i 902 volums, que podrien ser la meitat dels que van arribar a configurar la biblioteca en el moment del seu màxim esplendor amb el Seminari de Missions. Els antics posseïdors ens han donat noves pistes sobre la formació, l'estratigrafia de les col·leccions i les comunitats de lectors, que principalment han estat frares i clergues. La pedagogia devota va ser el centre d'atenció dels frares i al segle XVIII la Biblioteca es va convertir en un arsenal de llibres catòlics per servir de base dels tres eixos del convent: la formació de novicis, la preparació de la predicació i l'elaboració de textos devots.

Paraules clau: Patrimoni bibliogràfic, marques de propietat, exlibris, biblioteques, Escornalbou.

Las marcas de propiedad de los libros de la Biblioteca de Escornalbou. Un patrimonio recuperado

Resumen: La recuperación del patrimonio bibliográfico de Escornalbou se ha realizado sobre la base de los ex libris y *olim*; se han identificado 570 registros y 902 volúmenes, que pueden significar la mitad de la biblioteca en el momento de su mayor esplendor durante el Seminario de Misiones. Los anteriores poseedores nos han proporcionado nuevas pistas sobre la formación, la estratigrafía de las colecciones y las comunidades de lectores, que principalmente han sido monjes y clérigos. La pedagogía devota centró la atención de los frailes y en el siglo XVIII la biblioteca se convirtió en un arsenal de libros católicos que sirvieron de base a los tres ejes del convento: la formación de novicios, la preparación de la predicación y la elaboración de textos devotos.

Palabras clave: Patrimonio bibliográfico, marcas de propiedad, exlibris, bibliotecas, Escornalbou.

The ownership marks of the books of the Escorial Library: a recovered heritage

Resumen: The cultural heritage of Escorial was recovered on the basis of the exlibris and olim; 570 titles and 902 volumes have been identified, which could signify half of the library at its peak, with the Missions Seminary. Previous owners have given us some new pointers as to the formation, stratigraphy of the collections and the reader communities, comprised mainly of monks and the clergy. Devout pedagogy was the focal point of monks' attention, and in the 18th Century the library became an arsenal of Catholic books to act as the convent's triple cornerstone: the training of novices, the preparation of preaching and the production of religious texts.

Keywords: Cultural heritage, provenance, exlibris, libraries, Escorial.



1. La recuperació del patrimoni bibliogràfic més enllà de l'edició

La identificació de l'edició en el llibre antic ha suposat un dels reptes de la bibliografia i continua tenint un notable valor per analitzar impresos antics. Al mateix temps, l'estudi de l'exemplar està despertant molt d'interès, tal com posa de manifest David Pearson en el seu excel·lent llibre *Books as history: the importance of books beyond their texts*.¹ El llibre en ús, a través de les seves marques d'ús o propietat, revela que més enllà de l'edició hi ha nombroses preguntes sense respondre del pas dels llibres d'unes col·leccions a unes altres, així com dels usos dels llibres al llarg del temps. La traçabilitat dels exemplars ofereix perspectives innovadores d'estudi, tal com s'observa en el renovat interès per estudiar-ne les procedències. En el cas dels manuscrits, se'n pot seguir el pas a través de col·leccions: com en el projecte *Mapping Manuscript Migrations*, coordinat per Toby Burrows a la University of Oxford (<http://mappingmanuscriptmigrations.org/>) *Bibale* (<http://bibale.irht.cnrs.fr/>), que s'ocupa de la procedència dels manuscrits medievals de les col·leccions franceses, i disposa d'informació de 8.300 marques de procedència.²

L'estudi de les edicions a través de les marques d'antics posseïdors ha revelat col·leccions disperses, i de vegades, com en el cas de les actes de «Books and their owners» del Consortium of European Research Libraries³

i del projecte Material Evidence in Incunabula coordinat per Cristina Dondi, ha mostrat relacions insospitades entre exemplars.⁴ Aquests estudis revelen el procés de patrimonialització dels llibres, la mobilitat geogràfica i l'evolució de l'interès, al llarg del temps, per determinats materials convertits en objecte de col·lecció. Per exemple, l'ús de mètodes d'informació (MEI) permet seguir una edició des del seu lloc de producció fins a la biblioteca actual al llarg d'una línia del temps amb l'eina de visualització 15cV (<http://15cv.trade/>). La reconstrucció de biblioteques utilitzant les marques es combina amb els registres dels catàlegs, cosa que dona noves pistes sobre l'estratigrafia de les col·leccions patrimonials, que en el nostre àmbit es poden seguir en el cas de la base *Antics posseïdors* de la biblioteca de Reserva del Centre de Recursos per a l'Aprenentatge i la Investigació de la Universitat de Barcelona (CRAI-UB).⁵ Les biblioteques universitàries s'han afegit a aquestes iniciatives en comptar amb importants fons patrimonials i grups de recerca actius en l'estudi de les col·leccions, però les biblioteques públiques, especialment les provincials, tot i disposar d'un important fons patrimonial, no han iniciat pràcticament aquest tipus d'estudis.⁶ Sovint, els catàlegs, impresos o automatitzats, no informaven de les marques de propietat o els antics posseïdors s'anotaven sense normalitzar les dades dels exemplars, cosa que dificulta la recuperació dels catàlegs en línia.

En els estudis pioners del segle XIX, l'anàlisi de la possessió del llibre formava part dels valors afegits a l'edició i interessava a col·leccionistes

1. David Pearson, *Books as history: the importance of books beyond their texts*, London: British Library; New Castle: Oak Knoll Press, 2008.
2. Hanno Wijsman, «Bibale renouvelée: une base collective pour les données de provenance des bibliothèques françaises» [en línia], *Libraria: pour l'histoire des bibliothèques anciennes* (29/06/2018), <<https://libraria.hypotheses.org/637>> [Consulta: 31/08/2018].
3. *Books and their owners: provenance information and the European cultural heritage*. David J. Shaw (ed.). London: CERL, 2005.
4. Cristina Dondi, «'15cBooktrade': an evidence-based assessment and visualization of the distribution, sale and reception of printed books in the Renaissance» [en línia], *Gazette du livre medieval*, n. 60 (2013), p. 83-101, <<https://doi.org/10.3406/galim.2013.2035>> [Consulta: 26/09/2018].
5. Isabel Astals, Marina Ruiz, Neus Verger, «La base de dades 'Antics posseïdors' de la Biblioteca de Reserva de la Universitat de Barcelona» [en línia]. *BID: textos universitaris de biblioteconomia i documentació*, n. 24 (2010). <<http://bid.ub.edu/24/astals1.htm>> [Consulta: 26/09/2018].
6. Ana Santos Aramburo, Marta Torres Santo Domingo, «La Biblioteca Histórica de la Universidad Complutense: una primera aproximación a sus procedencias». En: *La memoria de los libros: estudios sobre la historia del escrito y de la lectura en Europa y América*, Salamanca: Instituto de Historia del Libro y de la Lectura, 2004, v. 2, p. 265-286.

La biblioteca reconstruïda a través de les marques permet entendre com es va formar la col·lecció, qui van ser els usuaris potencials (i reals) i quins usos va tenir, de tal manera que l'anàlisi de les anotacions forma part d'una visió més global de la història social i cultural del llibre.

i bibliòfils, però tal com ens adverteix Clavería «la historia del ejemplar y de sus poseedores anteriores es demasiado amplia para reducirla a la vanidad que proporcionan». ⁷ Allò que ens interessa és recalcar com és que les marques permeten reconstruir la col·lecció i donar una imatge de la comunitat de lectors interessada en una biblioteca. L'anàlisi dels fons patrimonials revela, en aquest camp, dades poc treballades dels llibres i els seus usos. En el nostre cas, intentarem incardinar els textos amb els seus usuaris i relacionar-los amb els objectius de l'orde franciscà en l'àmbit de la formació i la predicació, essent aquest últim aspecte el que permet conèixer la projecció dels textos en un territori, a través de les missions i de la publicació d'obres devotes. La biblioteca reconstruïda a través de les marques permet entendre com es va formar la col·lecció, qui van ser els usuaris potencials (i reals) i quins usos va tenir, de tal manera que l'anàlisi de les anotacions forma part d'una visió més global de la història social i cultural del llibre, en connexió amb algunes de les línies actuals d'estudi que proposa Raven. ⁸

Les biblioteques d'Escornalbou són plurals i se succeeixen al llarg dels segles, del segle XIII al XIX, però en tots els casos es tractaria

de biblioteques orientades a les necessitats de cada comunitat amb un marcat sentit pràctic. Un denominador comú que tenen és la temàtica religiosa que van posant al dia canonges, franciscans recol·lectes i franciscans observants, malgrat la manca de recursos. Amb l'arribada dels franciscans a finals del segle XVI, la biblioteca s'orienta a la predicació, cosa que va ser un dels exercicis comuns dels religiosos, i que s'incrementa amb el Seminari de Missions. El paper de la teologia moral es va enfortir incorporant les sumes morals que categoritzaven els pecats. L'exegesi bíblica va ser una font inesgotable de recursos per reforçar les al·legories devotes. D'aquesta manera, els llibres van ser l'arsenal de les missions i la lectura va ser un entrenament intel·lectual per a la predicació. Els textos que els religiosos elaboraven en el convent van esdevenir en bona part un recurs complementari per a predicar fent servir lletres d'impremta, com també ho feien en els seus viatges a les àrees rurals dels bisbats catalans, especialment els de Lleida, Tarragona i Tortosa, que van ser els més missionats. ⁹ Aquests textos es van nodrir dels abundants fons de la biblioteca. La formació, la predicació i l'elaboració de textos devocionals van ser els tres eixos al voltant dels quals els usuaris de la biblioteca van centrar els usos dels llibres. Uns llibres al servei del proselitisme catòlic.

De la biblioteca d'Escornalbou en conservem inventaris dels anys 1580-1584 i 1686, i un manuscrit anomenat *Index indicum* que indexa per matèries els llibres de la Biblioteca en el primer terç del segle XVIII. En el nostre cas, un primer inventari (1580-1584) ens apropa a una biblioteca que molt possiblement pertany, en part, als canonges de sant Agustí, tal com assenyala José Martín Mayor, historiador franciscà. ¹⁰ En el primer inventari, els guardians Pere Perelló i Miquel Seró van afegir més

7. Carlos Clavería, *Contra la bibliofilia: no amarás los libros sobre todas las cosas*, Madrid: Turpin Editores, 2015, p. 14.

8. James Raven, *What is the history of the book?*, Cambridge, UK; Malden, MA: Polity Press, 2018, p. 83-114.

9. Jaume Borràs Galceran, «Pensament i expressió religiosa: geografia i incidències històriques de les missions populars d'Escornalbou» [en línia], *Analecta sacra tarraconensia*, v. 67, n. 2 (1994), p. 95-96. <http://www.icatm.net/biblioteca/balmes/sites/default/files/public/analecta/AST_67_2/AST_67_2_93.pdf> [Consulta: 25/09/2018].

10. Josep Martí Mayor, «La biblioteca y archivo del Monasterio de Escornalbou», *Archivo Ibero-Americano*, v. 36, n. 142-143 (1976), p. 341-373.

títols a algunes seccions entre els anys 1584 i 1588, títols que en bona part són de temàtica franciscana. És interessant constatar que l'inventari documenta la presència d'autors clàssics com ara Aristòtil, Plató, Ciceró, Ovidi i Virgili, obres de literatura profana heretades de la Biblioteca dels canonges que són contràries al pensament recol·lecte centrat a atendre l'observança regular i orientat a perfeccionar la vida cristiana, tot allunyant-la d'interferències i perills innecessaris. A l'Escornalbou recol·lecte es documenta un intent de venda de llibres que pot considerar-se una mostra de la malfiança cap als clàssics.¹¹ El pare Pere Perelló s'oposà a la venda, fou un personatge destacat a la seva època dins de l'orde, amb una gran estima pels llibres, i va aconseguir l'objectiu de mantenir la Biblioteca intacta.

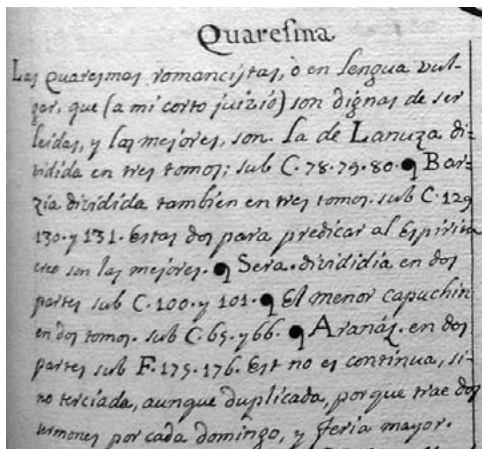


Figura 1. Sermons de quaresma de la biblioteca del convent d'Escornalbou, recomanats com a lectura pel pare Boada, amb les seves signatures per trobar els llibres a la biblioteca. Font: Francesc Boada Parellada, *Index indicum seu repertorium generale huius bibliothecae*, [Escornalbou, segle XVIII], Biblioteca Pública de Tarragona (BPT), ms. c. 1727-1728, f. 489r.

Més endavant, entre el 1727 i el 1728, Francesc Boada Parellada va elaborar l'*Index indicum seu repertorium generale huius bibliothecae*, un manuscrit que es va destinar a guiar els usu-

aris en les matèries de predicació, indicant en cada cas les referències als llibres disponibles a les prestatgeries, que eren els que podien llegir els frares;¹² alguns d'aquests s'han pogut catalogar com a part del fons patrimonial de la Biblioteca Pública de Tarragona. La possibilitat de seguir, a través d'aquests instruments de descripció, l'evolució de la col·lecció és un altre aspecte destacable del cas d'Escornalbou.

2. La biblioteca d'Escornalbou: formació, desfeta de la col·lecció i recuperació patrimonial

El lloc d'Escornalbou, documentat l'any 1198, acollí canonges de sant Agustí, antic orde dedicat al culte i a la predicació amb dependència de la mitra tarragonina; hi van viure prop de quatre segles en petites comunitats arrelades al territori, un fet que demostren deixes i donacions procedents del Camp de Tarragona. Durant el segle XIV entraren en una llarga decadència i a mitjan segle XVI només hi quedava un canonge. En vista d'aquesta situació, el que llavors era arquebisbe de Tarragona, Gaspar Cervantes de Gaeta (1511-1575), demanà al Papa la seva extinció, feta efectiva l'any 1574.



Figura 2. Panoràmica en picat del castell-monestir de Sant Miquel d'Escornalbou. 1916. Font: Arxiu Nacional de Catalunya, ANC1-132-N-325, Col·lecció fotogràfica d'Agustí Duran i Sanpere [en línia], <<http://arxiusenlania.cultura.gencat.cat/>>.

11. Eduard Toda, *Història d'Escornalbou*. Tarragona: Reial Societat Arqueològica, 1926, p. 161.

12. Jaume Bofarull, «Fra Francesc Boada, missionista d'Escornalbou i el seu manuscrit bibliogràficament notable 'Index indicum', composta a n'el 1727-28», *Butlletí Arqueològic*, època III, n. 39 (1929), p. 57-66; Agustí Boadas i Llavet, *Els Franciscans a Catalunya: història, convents i frades* (1214-2014), Barcelona: Província Franciscana de Catalunya; Lleida: Pagès, 2014, p. 183.

Amb la desaparició dels canonges, l'Arquebisbat hi va voler portar una comunitat de frares, que van trobar en els franciscans recol·lectes, que expandien llavors la seva reforma; aquests es mantingueren al convent fins l'any 1686, quan els observants els rellevaren per fer un Seminari de Missions. Eren comunitats petites i autosuficients que conreaven les terres i aprofitaven boscos i pastures. La predicació va ser important, més en períodes litúrgics com la quaresma, proporcionant-los ingressos. La lectura i la preparació de sermons ocuparen un lloc central, i la biblioteca heretada dels canonges experimentà un fort creixement al llarg del segle XVII, de cent-cinquanta fins a set-cents volums.



Figura 3. El convent d'Escornalbou. Detall de la portada del *Rosari Glossat* [Barcelona, 1928]. Font: Jaume Barrera i Escudero, *Sant Miquel d'Escornalbou*. 1921-1929, BC, Ms. 8995, f. 4.

L'Escornalbou recol·lecte va ser un lloc de referència espiritual al Camp i al Priorat que despertà grans mostres de religiositat i devoció cap als seus frares.¹³ El convent es nodria de religiosos de llocs propers, alguns de renom com ara el pare Pere Perelló (Cabacés, 1540-Escornalbou, 1602), guardià d'Escornalbou i definidor dels recol·lectes el 1585,¹⁴ i Bo-

naventura Gran (Riudoms, 1620-Roma, 1684), que va professar, l'any 1641, Escornalbou com a llec sense saber llegir;¹⁵ més tard anà a fundar el convent de Móra i la reforma del Sacro Ritiro a Itàlia.

A mitjan segle XVIII, i en el context del desplegament de les missions a Amèrica, l'any 1868 es fundà el Seminari de Missions, que perdurà fins a l'exclaustració el 1835. Si bé l'objectiu inicial era reclutar seminaristes per a les terres americanes, en van ser pocs al final, i la tasca evangelitzadora es concentrà en terres catalanoparlants, en les anomenades «missions populars», que es van repetir fins l'any 1833. Es tractava de predicacions extraordinàries fetes a les zones rurals, considerades manifestacions de la cultura barroca, properes a l'espectacle i al teatre.¹⁶ A Igual que els recol·lectes, els missioners no disposaven de rendes, vivien de la caritat, en comunitats nombroses, de quaranta o cinquanta persones.

La vida intel·lectual d'Escornalbou fou intensa, es varen formar predicadors i escriptors, i es van produir un gran nombre de textos de diversa índole. L'any 1765 el pare Joan Papió va descriure la Biblioteca com una de les més destacades de la província¹⁷ en relació amb la seva gran activitat centrada en la sagrada escriptura i moral. Va créixer dels set-cents als dos mil volums¹⁸ i disposà de bibliotecari des de 1734.¹⁹

L'any 1835, amb l'exclaustració, el Seminari d'Escornalbou es clausurà i els seus béns van ser expropiats per l'Estat, que els va gestionar amb pocs mitjans i una notable improvisació.

13. Eduard Toda, *Història...*, p. 99-101.

14. Ignasi Fernández Terricabras, «Unitat i diversitat en un orde religiós: les disputes entre franciscans observants i recol·lectes a Catalunya en temps de Felip II». *Pedralbes: revista d'història moderna*, n. 28 (2008), p. 363 i 369.

15. Josep Martí Mayor, «El ms. 4-001 de los franciscanos recoletos catalanes de Escornalbou (1580-1686)», *Archivo Ibero-Americano*, v. 51, n. 201-202 (1991), p. 190.

16. Francisco Rico Collado, «Conversión y persuasión en el barroco: propuestas para el estudio de las misiones en la España postridentina» [en línia]. *Studia Historica: historia Moderna*, v. 24, n. 1 (2002), p. 366, <http://revistas.usal.es/index.php/Studia_Historica/article/view/4720/4736> [Consulta: 25/09/2018].

17. Juan Papió, *El Colegio Seminario del Arcángel San Miguel de Escornalbou*, [Barcelona]: Imp. de los Carmelitas d., 1765, p. 47.

18. Agustí Boadas i Llavat, «Trovadores, filósofos y traductores: franciscanos catalanes a través de los siglos» [en línia]. En: Antonio Bueno García (coord.), *La labor de traducción de los franciscanos*, Madrid: Editorial Cisneros, 2013, p. 11, <<http://www.traducion-franciscanos.uva.es/archivos/Boadas.pdf>> [Consulta: 25/09/2018].

19. Agustí Boadas i Llavat, «Trovadores, filósofos y traductores:...», p. 10.

Aquest fet va comportar una de les més importants pèrdues de patrimoni de la història contemporània. A Escornalbou els frares van fugir amenaçats per la crema de convents a Reus el 22 de juliol; abans, però, van portar els béns als pobles veïns o els van ocultar entre les parets del convent, pensant a tornar. Es van endur, també, alguns dels llibres com a objectes personals, i ben segur que hi va haver vendes, espolis i sortides de llibres del país. El desembre de 1835 es documenten inventaris de llibres d'Escornalbou procedents de cases: a la de Miquel Ferrer, a l'Argentera, s'hi trobà una bona quantitat de llibres dins de coixineres i sàrries. Els mesos següents a l'exclaustració es van localitzar, es van treure de cases i amagatalls, es traslladaren a Reus i finalment a Tarragona. Tot plegat, marcat per una constant manca de recursos de tota mena: diners, locals i persones preparades: el comissionat Jacint Pla, de Reus, conegut com a «Xafarrucs», va ser destituït per males pràctiques; tampoc es podien pagar els trasllats dels llibres d'Escornalbou retinguts a Reus, i s'apuntà la possibilitat de vendre'n.²⁰ No sabem del cert què va succeir, però a finals de 1836 arribaren al seu destí, a Tarragona, a l'edifici del Pallol, a la *Sociedad Económica de Amigos del País*. Es tractava de llibres en mal estat perquè en cap cas amagatalls, magatzems i transports van tenir les mínimes condicions de conservació.

Més tard, el seu destí va ser la Biblioteca Provincial, creada l'any 1846, i associada llavors a l'Institut de segon ensenyament. Es nomenà bibliotecari Joaquín Caballero (1814-1887), franciscà exclaustrat que va formar part del Seminari d'Escornalbou entre el 1833 i el 1834. Marxà cap a Itàlia el 1835 i retornà a Tarragona l'any 1844 per fer de professor de religió a l'Ins-

titut de segon ensenyament. Era una persona molt culta i un destacat orador, va insistir al cap polític de la necessitat de crear una biblioteca provincial perquè a l'edifici del Pallol hi havia entre sis mil i vuit mil llibres en un deplorable estat d'abandonament. La seva dedicació professional al capdavant de la Biblioteca durant més de trenta anys va fer possible que a la seva mort, l'any 1887, es parlés de la Biblioteca Provincial de Tarragona com una de les millors pels seus còdexs, amb un fons que triplicaria la xifra inicial.²¹ Malgrat la migradesa de recursos, a Joaquim Caballero li devem la recuperació més important del patrimoni bibliogràfic de les comarques tarragonines del segle XIX.

4. Les marques de propietat dels llibres d'Escornalbou

Els exlibris són marques de propietat que s'anoten o s'incorporen als llibres i ens permeten identificar la pertinença a una col·lecció; ja des de l'edat mitjana s'acostumaven a fer en forma d'anotacions manuscrites als còdexs; es poden trobar a la guarda o al vers de la coberta, a la portadella o a la portada.²² La identificació del propietari és un procés que pot incloure també les marques de gestió bibliotecària (signatures) o d'enquadernació, que permeten detectar la incorporació a una col·lecció.²³

La col·lecció del convent d'Escornalbou s'ha seleccionat d'acord amb exlibris que n'han permès la identificació; s'han agrupat un total de 570 registres i 902 exemplars. Els llibres poden portar un o dos exlibris (a la portada i/o a la portadella i/o al full de guarda), i en alguns ca-

20. Eduard Toda, *Història...*, p. 262-272.

21. *Diario de Tarragona*, (25/11/1887), p. 2-4.

22. David Pearson, *Provenance research in book history: a handbook*, London: British Library; New Castle: Oak Knoll Press, 1994; M^a del Pilar Cuesta Domingo, «Ex-Libris en la Real Academia de la Historia: notas para un estudio», *Boletín ANABAD*, v. 65, n. 3 (2015), p. 149-208.

23. Raymond Clemens, Timothy Graham, «Assessing manuscript origin and provenance», En: *Introduction to Manuscript Studies*, Ithaca: Cornell University Press, 2007, p. 117-128.



sos, quan es tracta d'obres en volums, l'exlibris consta només al primer volum, o en alguns.²⁴ Llavors, un segon element d'identificació ha estat el lloc del llibre, on hi ha les dades de la localització de l'exemplar als prestatges: l'*olim* d'Escornalbou s'identifica per una lletra majúscula, escrita directament sobre el pergami, de color negre, acompanyada, sota, de números currens. Les lletres, que anaven de la A fins a la M, indicaven els armaris. Podem disposar d'altres informacions rellevants que ens reporten dades sobre els llibres i la seva història com són els exlibris dels antics posseïdors. Per incorporar un llibre a la biblioteca, el procediment era simple perquè es ratllava l'antic posseïdor i s'hi afegia l'exlibris del nou, que manifestava la nova pertinença, en aquest cas a la biblioteca d'Escornalbou.

Els exlibris d'Escornalbou estan manuscrits amb ploma o pinzell directament sobre l'ori-

ginal. Només de manera puntual s'informen dades com ara l'any d'incorporació. És a dir, si són llibres manlevats a altres convents o si han estat donats per caritat a canvi de misses per pregar per l'ànima del donant. Disposem d'alguns preu, de pensaments, d'indexacions, o d'altres que ens poden apropar també a patrons de lectura. Els exlibris van ser revisats per Eduard Toda (1855-1941), escriptor i bibliòfil, que va consultar de primera mà els fons conservats a la Biblioteca de Tarragona abans de 1926, data de publicació de la *Història d'Escornalbou*.²⁵ Dedicava un apartat a «La llibreria», en què en descriu la composició, els exlibris i les anotacions, un fet que ens ha permès comparar les dades publicades l'any 1926 amb les actuals. Tot i que durant la Guerra Civil Espanyola (1936-1939) els fons van patir trasllats, les dades aportades per Eduard Toda, 450 obres i 935 volums, s'aproximen a les recuperades, 570 registres i 902 volums.

24. Un exemple de diversos exlibris als volums de Beda, *Opera theologica, moralia, historica, philosophica, mathematica & rhetorica ... in Vetus et Novum Testamentum ... Coloniae Agrippinae: Apud Ioannem Wilhelmum Friessem juniorem, 1688*, ms. a port. «Es de la llibreria de S. Miquel de Escornalbou» (v. 1), «Es de la llibreria del Seminari de Escornalbou» (v. 3 i 5), «Es de Escornalbou» (v. 6 i 8)

25. Eduard Toda, *Història...*, p. 159-176.

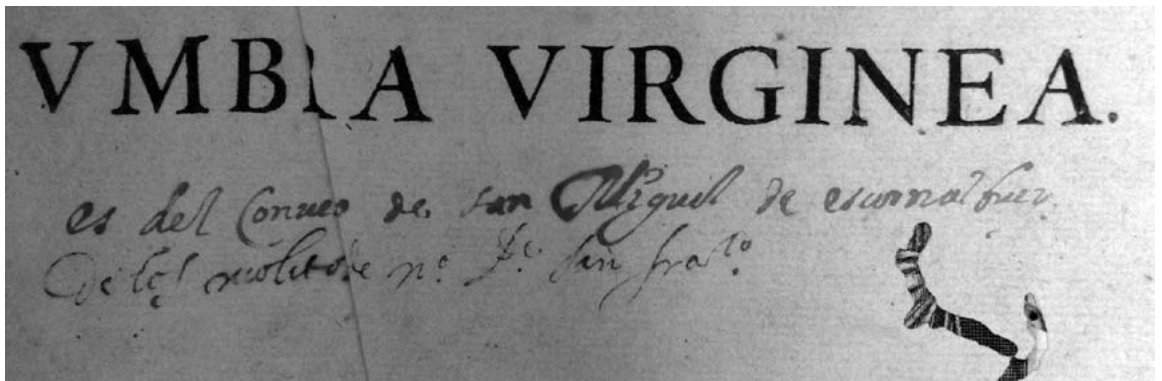


Figura 4. Un llibre de la biblioteca dels franciscans recol·lectes amb la marca de propietat a l'avantportada: «Es del Conuento de San Miguel de Escornalbuey de los recoletos de n[uestro] P[adr]e San Fra[ncis]co». Font: Luigi Novarini, *Electa sacra*. Turnoni [Tours]: sumpt. Laurentij Durand, & Laurentij Arnaud, 1640, BPT XI/479.

A continuació ens aturarem en les marques de propietat que manifesten la pertinença a les biblioteques d'Escornalbou, i en un segon punt, en els antics posseïdors, que ens han de permetre conèixer com es van formar les col·leccions.

4.1. Exlibris Scornalbou / Escornalbou

Per Eduard Toda, l'evolució dels exlibris va des del terme *Scornalbou*, atribuït als recol·lectes (segles XVI /XVII) al d'*Escornalbou*, emprat pels observants (segle XVIII), de lletra rodona i traç gruixut.

Comprovem que dels 903 volums només contenen el terme *Scornalbou* 36 títols, un 6,30%; a la resta, l'exlibris és «Escornalbou», topònim en català que és utilitzat majoritàriament d'ençà del segle XVI; només en els datats a les darreries del segle XV, en els incunables, representa un 50%. Per tant, és cert que l'exlibris amb el topònim *Scornalbou* s'usa més antigament i cau en desús el segle XVII, del qual hi ha set exemplars, i desapareix al XVIII, documentat només en un exlibris en llatí (Collegio Seminario de Scornalbou). Així mateix, hi ha variants que inclouen el topònim *Scornalbou*, la més comuna és *Llibreria de Sant Miquel de Scornalbou*,²⁶ fet que ens fa pensar que és una forma arcaï-

ca més pròpia de la comunitat agustiniana. El nombre de canonges, però, davalla a mitjan segle XVI i la seva col·lecció podria correspondre a part dels incunables, postincunables, i fins a les primeres dècades del segle XVI, ja que llavors Escornalbou queda pràcticament deshabitat. Sembla que els franciscans recol·lectes l'anaren abandonant perquè en l'inventari fet entre el 1581 i el 1584 ja es fa referència a la llibreria de Sant Miquel d'Escornalbou, denominació que es mantindrà també amb els observants del Seminari de Missions.



Figura 5. Exlibris manuscrit a la portada: «Es de la Libreria del con[ven]to de S[an]t Miquel de Scornalbou. 1665». Font: Enrique de Villalobos, *Summa de la theologia moral y canonica*, Valencia: por Silvestre de Esparsa a costa de Claudio Mace, 1630, BPT VIII-62.

Una tercera forma del mateix topònim és en castellà, *Escornalbuey*, la tenim només en dos llibres, l'un pertany a l'època dels recol·lectes i l'altre al Seminari.²⁷ Aquest darrer, però, té doble exlibris, i a més del terme en castellà n'hi consta un altre que diu: «És d'Escornalbou».

26. Llibreria Sant Miquel de Scornalbou (24 exemplars). Llibreria Scornalbou (4 exemplars). Comunitat de Scornalbou (4 exemplars). Saint Michaeli Scornalbou (2 exemplars). Llibreria del convent de Sant Miquel de Scornalbou (1 exemplar). Collegio Seminario Scornalbou (1 exemplar) (1716).

27. Manuscrit a portadella «Es del Conuento de San Miguel de Escornalbuey de los recoletos de n[uestro] P[adr]e San Fra[ncis]co» i ms. a f. de guarda «Es de la Libreria del S[an]to Seminario de Escornalbuey».

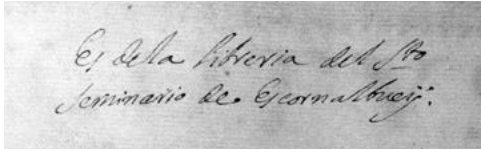


Figura 6. «Es de la Libreria del S[an]to Seminario de Escornalbuey». Font: Juan Francisco Pérez López, *Scotus philosophicus*, Barcinone [Barcelona]: ex thyp. Antonij Ferrer & Balthasaris Ferrer, 1687, BPT E2-35.

L'exlibris més emprat és Escornalbou («és d'Escornalbou»), i la seva forma més simple és la més utilitzada, amb prop d'un 50 % dels exlibris. En segon lloc, trobem el nom més complet, «llibreria d'Escornalbou» (16,3 %) i, amb més distància, «Seminari d'Escornalbou» (5,2 %). Altres exlibris són puntuals i no arriben en cap cas al 3 %. (Sant Miquel d'Escornalbou; Llibreria de Sant Miquel d'Escornalbou, Llibreria del Seminari d'Escornalbou, Santa Misión, Sant Miquel d'Escornalbou, Biblioteca d'Escornalbou, Convent de Sant Miquel d'Escornalbou i Comunitat d'Escornalbou).

4.2 Exlibris cancel·lats i la formació de les biblioteques d'Escornalbou

Els exlibris cancel·lats documenten el freqüent canvi de posseïdors en uns llibres amb anotacions, gastats i malmesos sovint per l'ús continuat. S'evidencia que se n'ha fet una lectura reiterada, sigui per preparar els sermons, per aprendre, per a la millora personal o la devoció. Els trobem en un 9 % de títols, tot i que cal tenir en compte que la xifra seria més elevada perquè les mutilacions, reenquadrernacions i el deteriorament han fet que perdem part d'aquesta informació.

Una primera pregunta que ens fem és: qui eren els antics posseïdors i/o usuaris? En primer lloc, trobem els llibres dels frares, que quan es moren queden al convent. Un primer exemple és el custodi fra Batista Ibanyes, expulsat del seu convent de València i refugiat a Escornalbou amb el seu *Blanquerna* de Ramon Llull, imprès a València el 1521; a la part superior de la portada hi consta que el deixa «per estar en esta casa». Una pràctica habitual als convents era que els frares rebessin els llibres en usdefruit havent de retornar-los a la comunitat conventual en marxar o en morir-se. En tenim vint exemples documentats que aporten la informació del frare: nom i cognom, amb la funció «ad usum» o en formes similars («usis concessim», «es para el simple uso», «es del uso simple», «és del simple ús...»); en pocs casos s'hi afegeixen altres informacions complementàries al costat del nom, com per exemple la condició del frare «jubilado», el segon cognom, el preu o una data: al segle XVII, Benet Soler, Francesc Simon Parareda. La primera part del segle XVIII, Josep Andreu, Joan Priu, Pere Bruges, i la segona meitat, Ludovic Pujol, Bernat Cavallé, Cabañas, Manuel Candano, Francesc Xavier Serra, Josep Manresa, Josefi Serret, Francesc Bordisses i Josefi Barbé Coll. Del XIX només hi trobem Rafael Blanc (consta l'ús de 1826, i la data d'edició és de 1804) i Miquel Piqué. En general, es tracta de llibres petits, un 68 % dels volums que hem documentat tenen menys de 22 cm, una mida que els fa idonis per al recolliment i la lectura personal. Un 26 % tenen entre 23 i 30 cm, i només el 10% tenen més de 30 cm. Pel que fa als temes, estan orientats a la predicació, a la devoció i, en menor grau, a la pràctica litúrgica, i molts corresponen a llibres editats la segona part del segle XVIII (63%), època d'intensa activitat al Seminari de Missions.

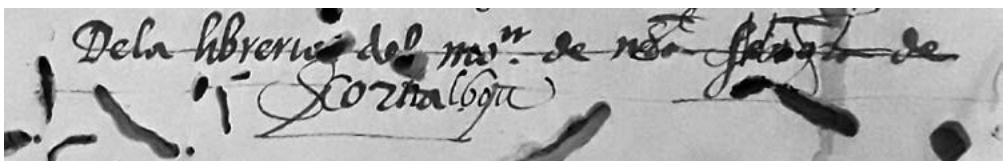


Figura 7. Exlibris manuscrit cancel·lat «De la libreria del mo[n]st[er]i de n[ost]ra Señora de Orta», i escrit a sobre «St. Miquel de Escornalbou». Font: *Sextus liber Decretalium per Bonifacium octauum*, Lugduni [Lió]: apud Hugonem à Porta & Antonium Vincentium, 1558-1559, BPT, XI-150.

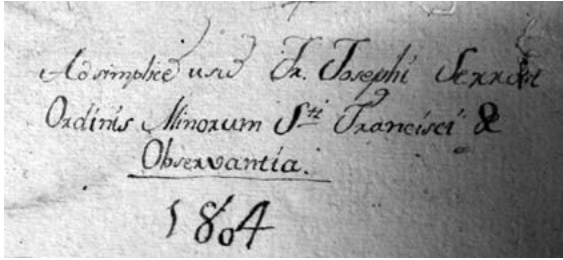


Figura 8. «Ad simplicem usu[m] Fr. Josephi Serrati Ordinis Minorum S[an]cti Francisci de Observantia. 1804». Font: Diego, de Estella, *Tratado de la vanidad del mundo*, En Madrid: en la imprenta de Joseph Otero, 1787, BPT IX-363.

Un segon grup són les donacions, que oscil·len des d'un o pocs llibres fins a deixes testamentàries més nombroses, com seria la que fa a darreries del segle XVIII Gabriel Munté, prevere i beneficiat d'Alforja, natural de la Baronia d'Escornalbou, que va reunir una important biblioteca per a l'època, 82 volums,²⁸ i que donà al Seminari de Missions d'Escornalbou a canvi que encomanessin la seva ànima a Déu. En conservem una part a l'actual Biblioteca.

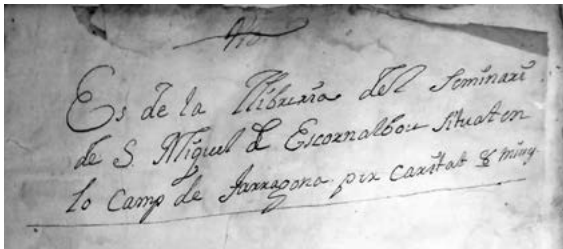


Figura 9. «Es de la libreria del Seminario de S. Miguel de Escornalbou situado en lo camp de Tarragona por caridad 8 missas». Font: Fèlix de Cabrera, *Catena moralis diversorum anulorum arte diligenter elaborata*, Cervariae [Cervera]: ex Typographiâ Reg. ac Pontificiae Universitatis per Thomam Senant, 1731, BPT E2-154.

Així mateix, la col·lecció actual testimonia la donació de Pere Miquel Aragonès, sacerdot de Cornudella, i se'n conserven cinc exemplars. Altres deixes correspondrien a rectors de pobles, preveres, vicaris i doctors. En general, eren eclesiàstics de comarques properes, de

28. Eduard Toda, *Història...*, p. 22-23.

29. Josep Martí Mayor. «El ms. 4-001...», p. 180.

30. Eduard Toda, *Història...*, p. 164-165.

Lleida i de Tortosa. Només en alguns consta la data de la donació, com és el cas d'Antoni Pallarès, que és de 1751. Observem, atenent-nos als anys d'impressió dels llibres, que el 84 % daten dels segles XVIII i XIX, cosa que podria tenir relació amb els itineraris de les missions fetes des del Seminari.



Figura 10. Exlibris manuscrit a portada i cancel·lat «Ex libris D[octo]ris Petri Michaelis Aragonis p[re]s[b]iter de Cornudella» i «Escornalbou». Font: Jaime, de Corella, *Practica de el confessorario*, En Madrid: en la imprenta de Juan Garcia Infançon: a costa de los herederos de Gabriel de Leon, vendese en su casa..., 1697-1698, BPT VIII/77.

Una tercera via d'entrada de fons són les compres, encara que són poc abundants n'hi ha que estan documentades; és el cas del pare Batista Ibanyes, que essent guardià del convent va comprar l'any 1595 la *Summa* de Rodrigues i llibres per al cor; en sabem també l'import, quinze lliures pagades pels Jurats de Montbrí, per la predicació d'una quaresma (dotze lliures), amb una donació de Joan Çabastida, conseller en cap de Barcelona (tres lliures).²⁹



Figura 11. «Attinet ad Bibliothecam Collegii Seminarii Missionario[rum] S. Michaelis de Escornalbou... emptus Romae 10 precio juliorum anno Domini 1722». Font: Jremlias Drexel, *Operum tomus secundus*, Lugduni [Lió]: sumptibus Ioannis-Antonii Huguetan, [1675], BPT XI/67.

De l'època del Seminari de Missions és interessant la compra de llibres a Roma que va fer el pare Francesc Boada, antic guardià que els portà a Escornalbou en retornar l'any 1724 després

de l'exili al qual l'obligà Felip V, en relació amb el posicionament d'Escornalbou a favor de l'arxiduc Carles d'Àustria. Des del convent de San Pietro in Montorio es dedicà a escriure cantorals i llibres, fet pel qual se'l va remunerar.³⁰ Contenen unes anotacions precises (lloc de compra, data, preu) escrites amb una excel·lent cal·ligrafia.

5. La recuperació: resultats obtinguts

L'anàlisi dels llibres del fons antic ha permès recuperar 570 registres bibliogràfics i 902 volums amb exlibris del monestir d'Escornalbou. Aquests llibres han estat catalogats per Núria Cangròs, Cristina Forn i Cristina Barbé en el marc de diverses campanyes durant els anys 2017 i 2018, incorporant-hi nous registres o les dades de l'exemplar al Catàleg del Patrimoni Bibliogràfic de Catalunya (CCPBC). Aquest primer pas ha permès identificar de visu els exemplars, verificar-ne l'edició i carregar les dades en el registre del CCPBC amb la finalitat que siguin accessibles per als investigadors. El resultat ha permès millorar la descripció de les edicions, a més d'ampliar de manera més detallada la informació dels exlibris en el registre de l'exemplar, transcrivint les anotacions de propietat i indicant «Fons Escornalbou» per a facilitar-ne la recuperació. En aquest procés, han quedat pendents quatre manuscrits i cinc volums impresos localitzats el mes de setembre de 2018 que no han estat incorporats encara en el CCPB.

Volem fer un advertiment previ que creiem que és necessari de cara a tractar amb prudència les dades que exposarem a continuació. A la part estadística ens referim a registres bibliogràfics del CCPBC, però en nombroses ocasions no són una obra, ja que, per exemple, trobem nombrosos sermonaris en diverses parts, textos en diversos toms, impresos en diferents impremtes, i fins i tot que procedeixen de diferents edicions, i en molts casos volums solts, fet habitual en les col·leccions desamortitzades.

L'anàlisi que fem es refereix només als llibres localitzats a la Biblioteca Pública de Tarragona amb exlibris que s'han conservat. És probable que durant les desamortitzacions del segle XIX hi hagués ocultacions i pèrdues d'obres d'interès, i algunes passessin a mans de col·leccionistes.

Els resultats de la taula 1 proporcionen una visió general del fons de la biblioteca conventual de notable interès: un 82 % dels registres són dels segles XVII i XVIII, coincidint amb el moment de més esplendor de predicació del Seminari de Missions; convé recordar que els frares van visitar, fins a la seva exclaustació de 1835 un total de 1.165 llocs en 3.184 campanyes de missions. Els llibres dels segles XV i XVI representen un petit percentatge, l'1,8%, un fet que pot respondre tant a pèrdues com també a la menor entitat de la biblioteca durant el segle XVI. En el cas dels incunables, són obres en llatí d'autors religiosos com Jean Gerson, sant Agustí i Giovanni Balbi, a més d'obres històriques com ara el *Fasciculus temporum omnes antiquorum cronicas complectens* de Werner Rolevinck. Els llibres del segle XIX, dels quals només hi ha 12 títols, són en bona part obres en castellà amb alguns textos en llatí o en italià, predominantment religiosos, de les quals la darrera que es va trobar és un llibre editat a Reus l'any 1837, *La verdadera esposa de Jesucristo* d'Alfonso Maria de Liguori, que porta un exlibris manuscrit a la portadella: «Fr. Miquel Pique religiós de Escornalbou». En aquest cas, es tracta d'un monjo exclaustat que va mantenir la connexió amb la comunitat d'origen.

Segle	Registres	%	Volums	%
XV	10	1,8	12	1,3
XVI	81	14,2	82	9,1
XVII	201	35,2	254	28,1
XVIII	266	46,6	528	58,5
XIX	12	2,3	26	3,0
Total	570	100	902	100

Taula 1. Llibres d'Escornalbou a la BPT per segles

A la taula 2 s'aprecia el pes del llatí (el 60 % dels registres) essent sempre superior el total de llibres en llengües clàssiques en tots els segles excepte al segle XIX —en llatí, és l'esplèndid exemplar amb gravats censurats de *Les metamorfosis* de Publi Ovidi Nasó de l'edició veneciana del 1509, però també trobem algun llibre bilingüe, en llatí i en grec, bíblies poliglotes o diccionaris multilingües. La poca presència d'obres en llengües vernacles els segles XV i XVI dona pas a un notable increment d'obres durant els segles XVII i XVIII, especialment els sermonaris, que van ser una peça essencial de les lectures: com ara els de José de Barcia, el *Despertador christiano de sermones doctrinales*; les nombroses obres devotes franciscanes com el *Libro de los tratados de Christo* (1630) de Rodrigo del Portillo, i les eines d'estudi com ara la *Rhetorica christiana* (1647), de Joan Baptista Escardó.

Llengua	Registres	%	s. XV	s. XVI	s. XVII	s. XVIII
Llengües vernacles	228	40		11	79	128
Llatí	342	60	10	70	122	138
Total	570	100	10	81	201	266

Taula 2. La distribució global del llatí i llengües vernacles a la biblioteca d'Escornalbou.

En qualsevol cas, la taula 3 mostra la important presència de llibres en castellà (37 %), seguida de molt lluny pel català (un escàs 1,4 % amb només vuit obres) i l'italià (1,2 %). El francès i el portuguès hi tenen una presència molt limitada. En el cas dels textos catalans, és sorprenent el fet que no es trobin les obres publicades per frares del convent, amb l'excepció de la *Font Mystica i sagrat del paradís de l'església* de Francesc Baucells. Una joia de la col·lecció és l'edició de 1521 de *Blanquerna* de Ramon Llull. Les obres pràctiques per a l'ús dels religiosos en les seves tasques evangèliques són els textos més comuns, com ara el *Promptuari moral sagrat y cathecisme pastoral de platicas doctrinals y espirituales sobre tots los punts de la doctrina christiana* (1754-15757) de Pere Salses i Trilles.

Llengua	Registres	%	s. XV	s. XVI	s. XVII	s. XVIII	s. XIX
Castellà	210	37		9	73	119	9
Català	8	1,4		2	4	2	
Francès	1	0,2				1	
Italià	7	1,2			1	5	1
Llatí	342	60	10	70	122	138	2
Portuguès	2	0,4			1	1	
Total	570	100	10	81	201	266	12

Taula 3. Les llengües dels llibres d'Escornalbou conservats a la BPT.

En la taula 4 s'observa el lloc d'impressió dels llibres del convent. Un total de 414 registres (el 72,3 % del total) es van imprimir en cinc ciutats (Barcelona, Madrid, Venècia, Lió, París i Roma); 148 registres es van imprimir en 38 ciutats, i en vuit casos es tracta d'impresos sense indicació de lloc d'impressió. El resultat és revelador d'una concentració d'alguns centres productors. Pel que fa a Catalunya, la preponderància de Barcelona és clara, de Tarragona hi ha sis impresos, de Cervera tres i de Vic hi ha únicament l'erudita biografia *Sanctus Severus, episcopus et martyr*, de Jaume Caresmar. El cas de Madrid és interessant, ja que ocupa el segon lloc, amb llibres fonamentalment del segle XVIII (63 impresos). No obstant això, en el cas de Barcelona van predominar els llibres del segle XVII (55), seguits dels de la centúria següent (46).

És interessant observar la diferència entre la producció realitzada a Espanya (amb un predomini de les llengües vernacles, especialment del castellà), amb un percentatge d'obres en

Lloc de publicació	Registres	Llatí	%	Llengües vernacles	%
Barcelona	106	26	25	80	75
Madrid	98	21	21	77	79
Venècia	92	89	97	3	3
Lió	71	71	100	-	-
París	25	24	96	1	4
Roma	22	19	86	3	14
Total	570	100	10	81	201

Taula 4. Principals lloc de publicació dels llibres d'Escornalbou conservats a la BPT.

Els estudis de la lectura necessiten informació contextual dels titulars i els lectors, que ens ha de permetre reconstruir les comunitats de lectors i les pràctiques culturals. El cas de la Biblioteca Pública de Tarragona és paradigmàtic de la recepció de fons valuosos de llibres arran de la desamortització.

llatí entorn del 20 %, en contrast amb els llibres importats de l'estranger, en els quals predomina el llatí, arribant en el cas de Lió al total dels registres. En el món eclesial, el llatí va mantenir el seu ús com a llengua de transmissió de saviesa, especialment en les obres erudites i els estudis teològics.

6. Conclusions

La història de les col·leccions, dels antics posseïdors i lectors ha ocupat un lloc en la historiografia del llibre, però és una tendència que, de ben segur, s'expandirà en els pròxims anys.³¹ En aquest estudi de cas, a la identificació de l'edició s'hi ha afegit el procés de reconstrucció de la trajectòria d'una col·lecció de llibres abans d'incorporar-se a la Biblioteca Pública de Tarragona. Aquest tipus d'estudis evidencien la necessitat de disposar de més bones eines de descripció i control, i mètodes d'anàlisi centrats en les anotacions i en la materialitat dels objectes.³²

31. Richard J. Kuhta, «Thys boke is myne: further reflections on the subject of provenance», *Gazette of the Grolier Club*, v. 54 (2003), p. 5-58.

32. Damien Blanchard, «Une archéologie du livre: les marques manuscrites comme source de l'histoire des bibliothèques bénédictines sous l'Ancien Régime». En: Bernard Dompnier et Marie-Hélène Froeschlé-Chopard (eds.), *Les religieux et leurs livres à l'époque moderne: actes du colloque de Marseille, E.H.E.S.S., 2 et 3 avril 1997*. Clermont-Ferrand: Presses universitaires Blaise-Pascal, 2000, p. 195-212.



Els estudis de la lectura necessiten informació contextual dels titulars i els lectors, que ens ha de permetre reconstruir les comunitats de lectors i les pràctiques culturals. En el cas de Catalunya, aquests estudis requereixen treballs detallats dels fons dels convents que van ser desamortitzats i llurs béns van passar a mans de biblioteques i d'arxius provincials. El cas de la Biblioteca Pública de Tarragona és paradigmàtic de la recepció de fons valuosos de llibres arran de la desamortització. En l'estudi de cas que proposem ens hem ocupat de la reconstrucció de la col·lecció del convent d'Escornalbou a partir d'una revisió de la catalogació dels llibres amb exlibris del convent, 570 registres amb 902 volums, que podrien ser la meitat dels que van arribar a configurar la biblioteca el segle XVIII, en el moment de més esplendor.

Les biblioteques d'Escornalbou són plurals i en tenim testimoni en els llibres conservats que pertanyen a les diferents èpoques i ens permeten apropar-nos als que foren els usuaris, encara que sigui de manera lleu per culpa de la seva llunyania. Uns trets comuns que tenen aquests llibres són la temàtica religiosa i el sentit pràctic, adaptats a les necessitats del moment i a l'època, i la fidelitat a l'orde religiós, en el sentit d'incorporar autors i textos.

Els testimonis més antics de l'època dels canonges de sant Agustí estan inclosos en el primer inventari (1581-1584), i són possiblement una bona part dels 170 volums heretats pels franciscans recol·lectes a la seva arribada a Escornalbou. És difícil, però, saber allò que formava part de la biblioteca dels canonges, possiblement els autors clàssics, textos sagrats i patristica. Podrien ser-ne part tant els incunables, amb l'exlibris en la seva forma arcaica (*Scornalbou*) com l'obra de sant Agustí, del seu orde: *Opuscula* (Venècia, 1484), o també Jean Gerson *Prima pars operum* (Nuremberg, 1489) o sant Bonaventura, *Prima-quarta pars in primum librum sententiarum* (1515).

L'arribada dels recol·lectes comporta l'adequació de la biblioteca a les noves necessitats, per això es manlleua el convent d'Horta de Sant Joan a Diego de Estella,³³ gran figura del franciscanisme hispà, i del qual s'incorporaran noves edicions a la biblioteca al llarg del temps. Ben aviat, s'hi inclou Àngel del Pas (1540-1596), impulsor de la reforma recol·lecta. Es manifesta la importància de l'art de predicar donant resposta a la recristianització del Concili de Trento, i augmenten els sermonaris.

Durant l'època del Seminari de Missions la biblioteca va ser la base dels tres eixos del convent: la formació de novicis, la preparació de les tasques evangèliques de la predicació i l'elaboració de textos devots. D'aquesta manera, al voltant de la lectura es varen dibuixar les biografies intel·lectuals dels religiosos. Els llibres llegits al convent que s'han conservat foren el 82 % dels exemplars dels segles XVII i XVIII, amb predomini del llatí, la llengua majoritària en 342 registres (60 % del total), seguit del castellà, un 37 % amb 210 registres. Els religiosos van transmetre aquests coneixements a les missions de predicació a través de l'oralitat dels sermons. Això va permetre que els llibres llegits al convent fossin filtrats a través dels interessos de la pedagogia devota i arribessin a tenir una presència continuada a les zones rurals que van visitar en campanyes missionals.

A la fi del segle XVII i durant el segle XVIII, la Biblioteca es convertirà en un arsenal de llibres catòlics. En aquest moment van formar part dels prestatges de la biblioteca algunes de les obres d'erudició bíblica de l'època, com la *Biblia maxima versionvm ex lingvis orientalibus* (1655-1660) i comentaris dels textos sagrats com els de Juan de Segovia i Cornelius à Lapide, les històries de l'església i l'orde franciscà i les obres hagiogràfiques, com les famoses *Acta sanctorum* de Johannes Bolland i Godefridus Henschenius. De la mateixa manera, es troben alineades amb l'ideari religiós algunes obres

33. Inventari anterior a 1584. Stela super Lucam, pars 1 i 2 (ja és al convent de Horta hont era), p. 352.

de teologia moral que oferien la categorització dels pecats, com el *Promptuari moral sagrat y cathecisme* (1754-1757) de Pere Salses i Trilles, i el *Directorio moral* (1786) de Francisco Echarri.

Entre els llibres conservats són pocs els casos de textos elaborats pels franciscans del convent, com ho va ser una edició de la *Font Mystica i sagrat del paradís de l'església* de Francesc Baucells. En altres col·leccions patrimonials, s'han conservat textos devots de gran difusió que es van editar a les impremtes de Reus, Tarragona i Girona, i que van afavorir la demanda de productes editorials senzills, de fàcil consum i destinats a un ús reiterat, ja que contenien oracions, guies de meditació, i altres recursos dirigits a persones etiquetades de «senzilles» o «rústiques», que d'aquesta manera tenien un conseller particular per resoldre els dubtes religiosos. La Biblioteca va ser el nucli en què es va centrar bona part de la comunitat de lectors del convent, cosa que va afavorir aquestes múltiples facetes culturals. La reconstrucció de la biblioteca ofereix, per primera vegada, de manera completa, una panoràmica dels llibres disponibles per als lectors de les comunitats que van ser usuàries d'aquests textos.

Bibliografia

ASTALS, Isabel; RUIZ, Marina; VERGER, Neus. «La base de dades 'Antics posseïdors' de la Biblioteca de Reserva de la Universitat de Barcelona» [en línia]. *BID: textos universitaris de biblioteconomia i documentació*, n. 24 (2010). <<http://bid.ub.edu/24/astals1.htm>> [Consulta: 26/09/2018].

BARBARÀ I CAMAFORT, Andreu. «El convent de Santa Anna». *Centre d'Estudis Alcoverencs: butlletí de divulgació*, n. 13 (1981), p. 17-25.

BARTOLOMÉ MARTÍNEZ, Bernabé. «Las bibliotecas públicas provinciales (1835-1885): un intento de promoción de la lectura en España» [en línia]. *Revista de educación*, n. 288 (1989), p. 271-304. <<http://www.mecd.gob.es/dctm/revista-de-educacion/articulosre288/re28812.pdf?documentId=0901e72b813ca29f>> [Consulta: 27/09/2018].

BLANCHARD, Damien. «Une archéologie du livre: les marques manuscrites comme source de l'histoire des bibliothèques bénédictines sous l'Ancien Régime». En: DOMPIER, Bernard; FROESCHLÉ-CHOPARD, Marie-Hélène (eds.). *Les Religieux et leurs livres à l'époque moderne: actes du colloque de Marseille, E.H.E.S.S., 2 et 3 avril 1997*. Clermont-Ferrand: Presses universitaires Blaise-Pascal, 2000, p. 195-212.

BOADAS I LLAWAT, Agustí. «Trovadores, filósofos y traductores: franciscanos catalanes a través de los siglos» [en línia]. En: BUENO GARCÍA, Antonio (coord.), *La labor de traducción de los franciscanos*. Madrid: Editorial Cisneros, 2013. <<http://www.traduccion-franciscanos.uva.es/archivos/Boadas.pdf>> [Consulta: 25/09/2018].

— *Els Franciscans a Catalunya: història, convents i frades* (1214-2014). Barcelona: Província Franciscana de Catalunya; Lleida: Pagès, 2014.

BOFARULL, Jaume. «Fra Francesc Boada, missionista d'Escornalbou i el seu manuscrit bibliogràficament notable 'Index ludicum', compost a n'el 1727-28». *Butlletí Arqueològic*, època III, n. 39 (1929), p. 57-66.

Books and their owners: provenance information and the European cultural heritage. David J. Shaw (ed.). London: CERL, 2005.

BORRÁS I GALCERAN, Jaume. «Pensament i expressió religiosa: geografia i incidències històriques de les missions populars d'Escornalbou» [en línia]. *Analecta Sacra Tarraconensia*, v. 67, n. 2 (1994), p. 93-114. <http://www.icatm.net/biblioteca/balmes/sites/default/files/public/analecta/AST_67.2/AST_67_2_93.pdf> [Consulta: 25/09/2018].

CLAVERIA, Carlos. *Contra la bibliofília: no amaràs los libros sobre todas las cosas*. Madrid: Turpin Editores, 2015.

CLEMENS, Raymond; GRAHAM, Timothy. «Assessing manuscript origin and provenance». En: *Introduction to Manuscript Studies*. Ithaca: Cornell University Press, 2007, p. 117-128.

CUESTA DOMINGO, M^a del Pilar. «Ex-Libris en la Real Academia de la Historia: notas para un estudio». *Boletín ANABAD*, v. 65, n. 3 (2015), p. 149-208.

DONDI, Cristina. «'15cBooktrade': an evidence-based assessment and visualization of the distribution, sale and reception of printed books in the Renaissance» [en línia]. *Gazette du livre medieval*, n. 60 (2013), p. 83-101. <<https://doi.org/10.3406/galim.2013.2035>> [Consulta: 26/09/2018].

FERNÁNDEZ TERRICABRAS, Ignasi. «Unitat i diversitat en un orde religiós: les disputes entre franciscans observants i recol·lectes a Catalunya en temps de Felip II». *Pedralbes: revista d'història moderna*, n. 28 (2008), p. 355-376. <<https://www.raco.cat/index.php/Pedralbes/article/view/234177>> [Consulta: 26/09/2018].

KUHTA, Richard J. «Thys boke is myne: further reflections on the subject of provenance». *Gazette of the Grolier Club*, v. 54 (2003), p. 5-58.

MARTÍ MAYOR, Josep. «El ms. 4-001 de los franciscanos recoletos catalanes de Escornalbou (1580-1686)». *Archivo Ibero-Americano*, v. 51, n. 201-202 (1991), p. 177-200.

— «La biblioteca y archivo del Monasterio de Escornalbou». *Archivo Ibero-Americano*, v. 36, n. 142-143 (1976), p. 341-373.

— «Escornalbou: colegio-seminario de misiones de Propaganda Fide: 1686-1835». *Archivo Ibero-Americano*, v. 42, n. 165-168 (1982), p. 293-342.

MARTÍN ABAD, Julián. «Del ejemplar, o de ciertas historias (casi) siempre pasadas por alto». En: *El enredijo de mil y un diablos*:

de manuscrits, incunables y raros, y de fondos y fantasmas bibliográficos. Madrid: Ollero y Ramos, 2007, p. 377-396.

MOLINER PRADA, Antoni. «El anticlericalismo popular durante el Bienio, 1834-1835» [en línia]. *Hispania Sacra*, v. 49, n. 100 (1997), p. 497-541. <<https://doi.org/10.3989/hs.1997.v49.i100.646>> [Consulta: 29/09/2018].

MUÑOZ FELIU, Miguel C. *Bibliotecas y desamortización en la ciudad de Valencia: (1812-1844)*. València: Universitat de València, 2015. <<https://dialnet.unirioja.es/servlet/tesis?codigo=73377>> [Consulta: 29/09/2018].

PEARSON, David. *Books as history: the importance of books beyond their texts*. London: British Library; New Castle: Oak Knoll Press, 2008.

— *Provenance research in book history: a handbook*. London: British Library; New Castle: Oak Knoll Press, 1994.

PONS ANGUERA, Antoni. *Libro de varias cosas sucedidas en esta villa y algunos parages de Cataluña*. Reus: Associació d'Estudis Reusencs, 1988.

RAVEN, James. *What is the history of the book?* Cambridge, UK; Malden, MA: Polity Press, 2018.

RICO COLLADO, Francisco. «Conversión y persuasión en el barroco: propuestas para el estudio de las misiones en la España postr-

dentina» [en línia]. *Studia Historica: Historia Moderna*, v. 24, n. 1 (2002), p. 363-386. <http://revistas.usal.es/index.php/Studia_Historica/article/view/4720/4736> [Consulta: 25/09/2018].

SABATÉ I BOSCH, Josep M. *Monges, frares, canonges, capellans, i capellansets: Tarragona religiosa a les acaballes de l'Antic Règim*. Tarragona: Ajuntament de Tarragona, 1992.

SANTOS ARAMBURO, Ana; TORRES SANTO DOMINGO, Marta. «La Biblioteca Histórica de la Universidad Complutense: una primera aproximación a sus procedencias». En: *La memoria de los libros: estudios sobre la historia del escrito y de la lectura en Europa y América*. Salamanca: Instituto de Historia del Libro y de la Lectura, 2004, v. 2, p. 265-286.

TODA, Eduard. *Història d'Escornalbou*. Tarragona: Reyal Societat Arqueològica, 1926.

— *Los Convents de Reus y sa destrucció en 1835: Francescans de Santa Maria de Jesús, Carmelitas de Sant Joan, Paülis del Seminari, Monjas Carmelitas*. Reus: Revista del Centre de Lectura, 1930.

WUSMAN, Hanno. «Bibale renouvelée: une base collective pour les données de provenance des bibliothèques françaises» [en línia]. *Libraria: pour l'histoire des bibliothèques anciennes*, (29/06/2018). <<https://libraria.hypotheses.org/637>> [Consulta: 31/08/2018]. ■



LA
TRANQUIL·LITAT
QUE DONA
ESTAR EN
BONES
MANS

IMPRESSIÓ OFFSET & DIGITAL
Informa-te'n a: www.sprintcopy.com



Sprint Copy
934 463 900

WE PRINT THE DIFFERENCE
Còrsega 546 - Tel. 934 463 900
sp@sprintcopy.com
www.sprintcopy.com