

INTEGRACIÓ D'ARXIUS, MUSEUS I BIBLIOTEQUES AMB TECNOLOGIES DEL WEB SEMÀNTIC: DE LA MODELITZACIÓ CONCEPTUAL A LA INICIATIVA LINKED OPEN DATA

Julio Quílez Mata

Director de l'Arxiu Comarcal
de l'Alt Urgell

1. INTRODUCCIÓ

Els darrers vint anys s'ha experimentat, en el camp de les tecnologies de la informació i de la gestió del coneixement, una transformació semblant a l'evolució que ha viscut la geometria euclidiana des de la irrupció de la geometria fractal a principis de la dècada del 1980. Tot tendeix cap a l'explicitació i la modelització de la desagregació per poder gestionar amb més facilitat sistemes complexos en entorns cada vegada més plàstics i dinàmics, pràcticament líquids. La rigidesa dels documents sembla superada per la textura plàstica i polimòrfica de les dades.

Una aproximació a l'evolució de les eines tecnològiques i la seva significació —que no implica necessàriament exclusió entre si— aporta una pista de la tran-

sició cap a la gestió d'entorns tecnològics cada vegada més oberts, transversals i complexos: del protocol d'intercanvi d'informació OAI-PHM s'ha evolucionat cap a l'OAI-ORE, que permet gestionar informació i objectes compostos i complexos; d'estructures taxonòmiques s'ha passat a ontologies, més sofisticades, més explícites, més profundes; de repositoris de documents (model docucèntric) s'està evolucionant cap a repositoris *triplestores* o de triplets RDF, és a dir, de dades i de semàntica (models datacèntrics); del web hipertextual al web de dades (computacional), fonament dels processaments *big data*; de models de metadades basats en binomis atribut-valor a models «tridimensionals» basats en ontologies orientades a esdeveniments que es modelitzen sobre el temps i l'espai; de normes i estàndards textuais i unitaris a representacions conceptuals que defineixen i expliciten components i comportaments intercanviables; de representacions del coneixement parcials a integració d'escenaris; de bases de dades relacionals a bases de dades NoSQL, i d'informació captiva en sitges estructurals i tecnològiques a dades obertes, vinculades, amb significat i reutilitzables, processables per màquines.

En aquest article s'aborda una aproximació succinta a diferents estratègies d'integració de les LAM (*libraries, archives and museums*), biblioteques, arxius i museus des de la perspectiva del web semàntic que incardina el doble vessant de la modelització del coneixement com a pas previ al desplegament de la tecnologia.

2. LA MODELITZACIÓ CONCEPTUAL COM A PAS PER A LA DEFINICIÓ I LA INTEGRACIÓ DEL CONEIXEMENT

Des de finals dels anys noranta del segle passat, davant les possibilitats que oferia el web des de la perspectiva de l'intercanvi d'informació, així com del creixement exponencial de models de metadades que el limitaven, primer des de la comunitat dels museus i posteriorment des de la de les biblioteques i els arxius, s'inicià un procés de disseny de marcs conceptuals per facilitar la integració del coneixement i el desenvolupament d'aplicacions.¹

2.1 UNA ONTOLOGIA DEL DOMINI DELS MUSEUS OBERT A TOT EL PATRIMONI CULTURAL: CIDOC CRM – ISO 21127

Es tracta d'un model conceptual desenvolupat pel Grup de Normalització Documental del Consell Internacional dels Museus (ICOM-CIDOC) i, posteriorment, per un grup de treball especialitzat al si de CIDOC, el CRM-SIG (Conceptual

Reference Model-Special Interest Group). Els treballs es van iniciar l'any 1994 i la primera versió de l'any 1996 va ser presentada seguint la metodologia del model entitat-relació. Per atorgar més expressivitat al model, es va decidir redisenyar-lo en forma de modelització orientada a objecte i, finalment, l'any 2006 l'ontologia va ser convertida a estàndard internacional per la ISO (International Standard Organization) com a ISO 21127:2006² i revisada posteriorment l'any 2014.³ La versió actual de la CIDOC CRM és la 6.2.2.⁴

Es tracta d'una de les eines fonamentals en molts dels projectes de modelatge de dades en el web semàntic per la seva riquesa i nivell d'estructuració. L'objectiu de l'ontologia és crear un marc de referència comú a fi que fonts d'informació heterogènies i incompatibles puguin alinear-se i harmonitzar-se i, en darrer terme, puguin transformar recursos d'àmbit local en un recurs universal coherent. Es tracta, per tant, de facilitar l'intercanvi i la recuperació d'informació aplicats al domini del patrimoni cultural.

El model identifica i defineix 86 conceptes o classes amb un gran nivell de detall, així com les seves relacions, que s'expliciten a través de 137 propietats igualment detallades en l'abast, el significat i l'aplicació. Es tracta, per tant, d'un model semàntic que permet identificar el significat explícit d'allò que és implícit i facilitar-ne l'estructuració lògica. D'altra banda, la manera com es relacionen les entitats per mitjà de les propietats produeix declaracions semblants a les gramaticals en forma de subjecte (una entitat) – predicat (la relació) – objecte (l'altra entitat), fet que permet una implementació fàcil en llenguatges ontològics com ara Resource Description Framework (RDF), RDF Schema (RDFS) i Ontology Web Language (OWL).

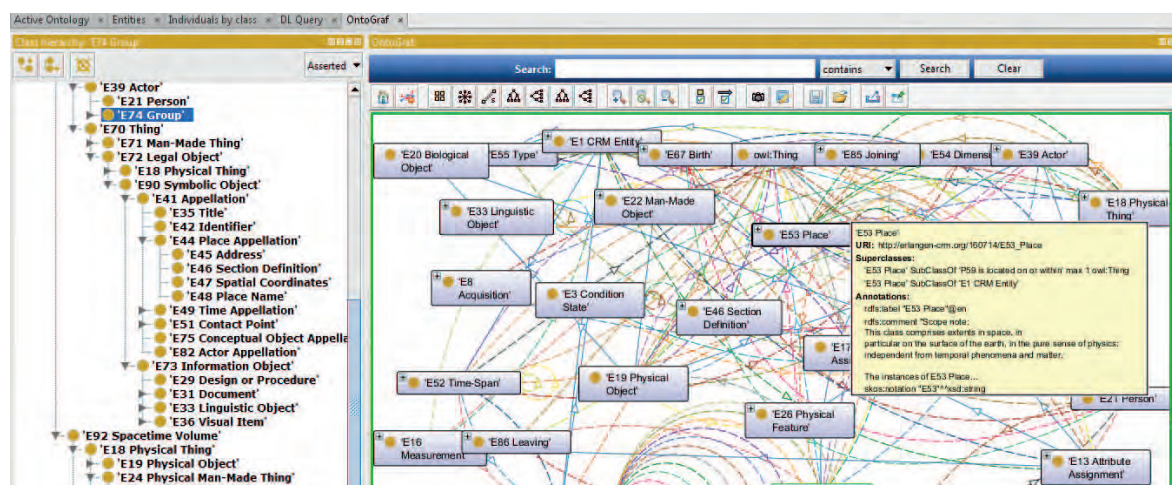


Figura 1. Visualització de l'ontologia CIDOC CRM mitjançant l'editor d'ontologies Protégé (versió 5.1.0)⁵ formalitzada en OWL en el context del projecte ERLANGEN.⁶

Els àmbits d'aplicació principals giren al voltant de la constitució d'una base de coneixement comuna entre professionals d'àmbits i perfils diferents durant les fases de disseny i implementació d'un programari informàtic; aporten un marc conceptual ja sigui per al desenvolupament de nous sistemes d'informació o per a l'avaluació dels existents al si de projectes de reenginyeria, i, finalment, possibiliten la federació de fonts d'informació heterogènies i permeten la recuperació de recursos des d'una única interfície sense una integració tècnica d'aplicacions, és a dir, a través de la construcció dels anomenats *sistemes de mediació*. És aquesta darrera aplicació la que situa el model en el pilar de molts desenvolupaments del web semàntic, especialment en el cas dels portals semàntics.

Probablement, un dels aspectes més rellevants de l'ontologia és la seva condició de model orientat a esdeveniments o *event centric*, a diferència d'altres esquemes destinats a la descripció de documents o *document centric*, com podria ser el cas de l'esquema Dublin Core. Aquesta orientació confereix al model la consideració de *semantic glue* (Hyvönen, 2012, p. 44) o cola semàntica, és a dir, que atorga una gran consistència i expressivitat semàntica als elements descrits, expressats o alineats en forma d'instàncies de les classes de l'ontologia per mitjà de la relació entre objectes, actors, temps, espai i esdeveniments. Això implica que les metadades deixen de descriure simplement documents i passen a formar part d'una constel·lació de coneixement molt més enriquida i complexa. Es podria dir que CIDOC CRM confereix una perspectiva tridimensional a l'expressivitat semàntica de les metadades en comparació dels models tradicionals, més plans. En aquest sentit, el model CIDOC CRM entra a formar un grup d'ontologies orientades a la gestió d'esdeveniments com ara ABC Ontology (Lagoze, Hunter, 2001), Event Ontology,⁷ DSNotify Eventsets,⁸ LODE,⁹ DOLCE-Dns Ultralite (dul)¹⁰ i OpenCYC Ontology¹¹ i a relacionar-s'hi.

És per aquest motiu que l'ontologia CIDOC CRM ha estat aplicada i s'aplica profusament en processos d'interoperabilitat entre sistemes per mitjà de l'harmonització d'esquemes de metadades del camp del patrimoni documental en general. Els promotors de CIDOC CRM, conscients de la necessitat d'integrar el coneixement a través d'una bona gestió tecnològica dels mapatges, l'any 2012 impulsaren la creació d'un model tècnic de referència¹² i, en aquests moments, centenars d'institucions i projectes ja han fet les seves correspondències i alineaments.¹³

L'estàndard també s'ha destinat a la creació i l'aplicació de regles semàntiques per a la inferència de coneixement per mitjà de l'extensió de l'ontologia a CRM-

dif Argumentation Model,¹⁴ i, també, s'ha utilitzat a bastament en el procés de transformació d'eines de control de vocabulari com ara tesaurus, classificacions o encapçalaments de matèries a la tecnologia *linked data* (LD) o dades enllaçades i *linked open data* (LOD) o dades obertes enllaçades. Un cas rellevant d'aplicació el constitueixen els tesaurus i els vocabularis proveïts pel Getty Research Institute, especialment en el cas de CONA (Cultural Objects Name Authority) (Harpring, 2014, p. 23).

Finalment, cal destacar la flexibilitat de l'ontologia per mitjà de la seva formulació telescòpica, és a dir, la possibilitat de la seva representació mínima en forma de taxonomia constituïda per una relació molt reduïda de classes o bé la capacitat de constituir extensions del model genèric que permet integrar altres comunitats professionals més enllà de la museística i del patrimoni cultural en general. Actualment hi ha fins a vuit especialitzacions: CRMInf Argumentation Model,¹⁵ CRMarchaeo Excavation Model,¹⁶ CRMsci Scientific Observation Model,¹⁷ CRMgeo Spatiotemporal Model,¹⁸ CRMdig Model for Provenance Metadata,¹⁹ CRMba Model for Archaeological Buildings,²⁰ FRBRoo Functional Requirements for Bibliographic Records²¹ i PRESSoo Modeling of Bibliographic Information.²²

2.2 MODELITZACIONS CONCEPTUALS EN EL CAMP DE LES BIBLIOTEQUES

L'any 1992, la IFLA (International Federation of Library Associations and Institutions) inicià un procés de modelització conceptual que va finalitzar l'any 2010 amb tres ontologies: FRBR (Functional Requirements for Bibliographic Records),²³ presentada l'any 1997 sobre els requisits funcionals dels registres bibliogràfics i dues extensions d'aquesta ontologia, una relativa a les dades d'autoritat anomenada FRAD²⁴ (Functional Requirements for Authority Data), finalitzada el 2009, i l'altra, FRSAD (Functional Requirements for Subject Authority Data), orientada a les relacions de matèria, publicada l'any 2010. Recentment, a finals de febrer del 2016, la IFLA ha presentat per a la seva revisió una proposta d'unificació dels tres models conceptuals en un de sol amb el títol de FRBR-LRM (Library Reference Model),²⁵ el qual té per objectiu constituir un model conceptual de referència d'alt nivell.

El model FRBR està conceptualitzat mitjançant el modelatge entitat-relació i analitza el registre catalogràfic com un conjunt d'informacions sobre quatre aspectes diferenciats del document descrit mitjançant característiques relatives a

l'exemplar (ítem), a la publicació (manifestació), al contingut intel·lectual o artístic (expressió) i a la creació abstracta (obra). El model consolida les pràctiques tradicionals de catalogació internacionals per mitjà de la seva racionalització i conceptualització amb la creació d'un marc conceptual i terminològic orientat al desenvolupament i l'adaptació de catàlegs innovadors d'acord amb el web semàntic i l'intercanvi d'informació mitjançant la iniciativa LD, que a la pràctica es manifesten en el model de catalogació internacional RDA (Resource Description and Access), concebut per reemplaçar les AACR (Anglo American Cataloguing Rules) (Picco, Ortiz, 2012), i BIBFRAME²⁶ (Bibliographic Framework), model de dades de descripció bibliogràfica dissenyat per substituir els formats MARC.²⁷

L'any 2003 es va dur a terme una transformació del model FRBR com una extensió de CIDOC CRM, per afavorir la interoperabilitat entre recursos procedents de biblioteques i museus. El nou model es va desenvolupar d'acord amb els principis de modelatge orientat a objectes, propi de CIDOC CRM, per afavorir la seva conversió al metamodel RDF i permetre integrar-lo amb les tecnologies de LD o LOD. A partir d'aquell moment, el model FRBR original va passar a ser designat, de manera no oficial, com a FRBRer (FRBR entitat-relació) i el resultat de la integració amb CIDOC CRM es va anomenar FRBRoo (FRBR orientat a objecte). Posteriorment, l'any 2012, una nova versió de FRBRoo 2.0 integrava en la modelització els models FRAD i FRSAD. L'extensió de FRBR ha permès al model gaudir de més expressivitat relativa a la gestió dels esdeveniments en la descripció dels recursos (Le Boeuf, 2013).

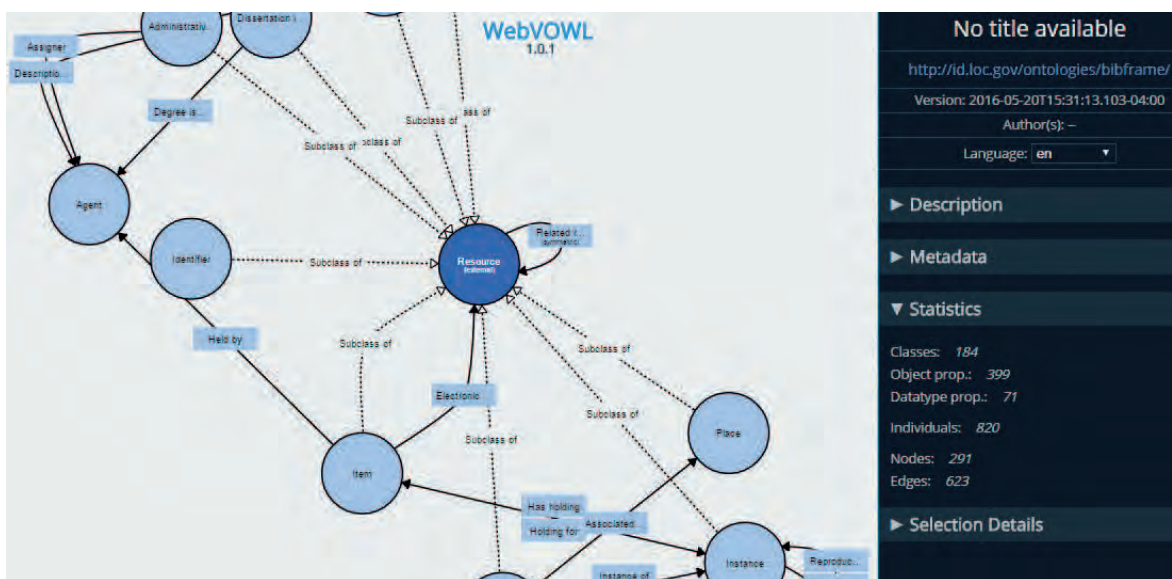


Figura 2. Representació gràfica d'una secció de l'ontologia BIBFRAME utilitzant el visualitzador en línia d'ontologies WebVOWL (versió 1.0.1),²⁸ carregada a partir del seu URI (*uniform resource identifier*) o IRI (*internationalized resource identifier*).

2.3 MODELS CONCEPTUALS EN EL DOMINI DE L'ARXIVÍSTICA

L'any 1989 marca l'inici del desenvolupament per part de l'ICA (International Council on Archives) de normes de descripció arxivístiques que, de manera gradual, es van anar publicant entre el 1994 i el 2008: ISAD (G), ISAAR (CPF), ISDF i ISDIAH. La manca de directrius sobre la seva aplicació i integració ha generat problemes d'implementació en sistemes d'informació i, a la pràctica, ha provocat la tendència a una profusa elaboració de descripcions unitàries que no separen la descripció arxivística del control d'autoritats, fet que ha limitat extremadament l'intercanvi d'informació entre arxius que comparteixen productors i sèries documentals en un moment en què les tecnologies de la informació permeten, mitjançant tecnologies semàntiques, integrar de manera global l'accés a enormes quantitats d'informació d'arxius, biblioteques i museus. La manca, per tant, d'una estandardització de les normes arxivístiques i d'un model conceptual semblant als que ja hi ha en altres comunitats com ara el CIDOC CRM en museus i la FRBR en biblioteques ha provocat, per part de diverses comunitats arxivístiques d'arreu del món, la necessitat de la modelització d'una visió integrada dels diferents elements que intervenen en els processos de creació, gestió i manteniment de documentació d'arxiu. Així, entre els models conceptuals arxivístics més veterans es troben els elaborats pels Arxius de Nova Zelanda publicats l'any 2008, l'Australian Government Recordkeeping Metadata Standard i les Technical Specifications for the Electronic Recordkeeping Metadata Standard, i, a l'altre extrem, se situa el Finnish Conceptual Model for Archival Description, publicat pels Arxius Nacionals de Finlàndia l'any 2013 com a versió de revisió.²⁹ Entremig n'han aparegut d'altres com ara les Normas Españolas de Descripción Archivística (NEDA), LOCAH a Anglaterra i el sistema AtoM patrocinat per l'ICA. Els models tenen en comú, a grans trets, la identificació d'entitats com ara document, agent, funcions, mandats i esdeveniments, com també la definició d'un seguit de relacions entre si. Dels diversos models, LOCAH, des d'un enfocament basat en l'EAD (Encoded Archival Description), i AtoM s'orienten vers la publicació dels registres com a *linked open data*.

En el cas espanyol, la NEDA-I, elaborada per la Comisión de Normas Españolas de Descripción Archivística (CNDA), s'ha definit per mitjà de la metodologia de modelatge entitat-relació. Els primers treballs s'iniciaren l'any 2007³⁰ i la primera publicació relativa a les entitats i les relacions es presentà l'any 2012.³¹ Recentment, l'any 2016, ha estat completada, en versió de revisió, per la destinada a la definició dels atributs³² com a propietats rellevants de les entitats o relacions del model. En conjunt, es tracta d'un treball molt elaborat partint d'una anàlisi molt

completa de la resta de normes conceptuals i estàndards arxivístics nacionals i internacionals i té en compte, a més, la totalitat de les normes de descripció arxivístiques autonòmiques, entre les quals hi ha la NODAC.

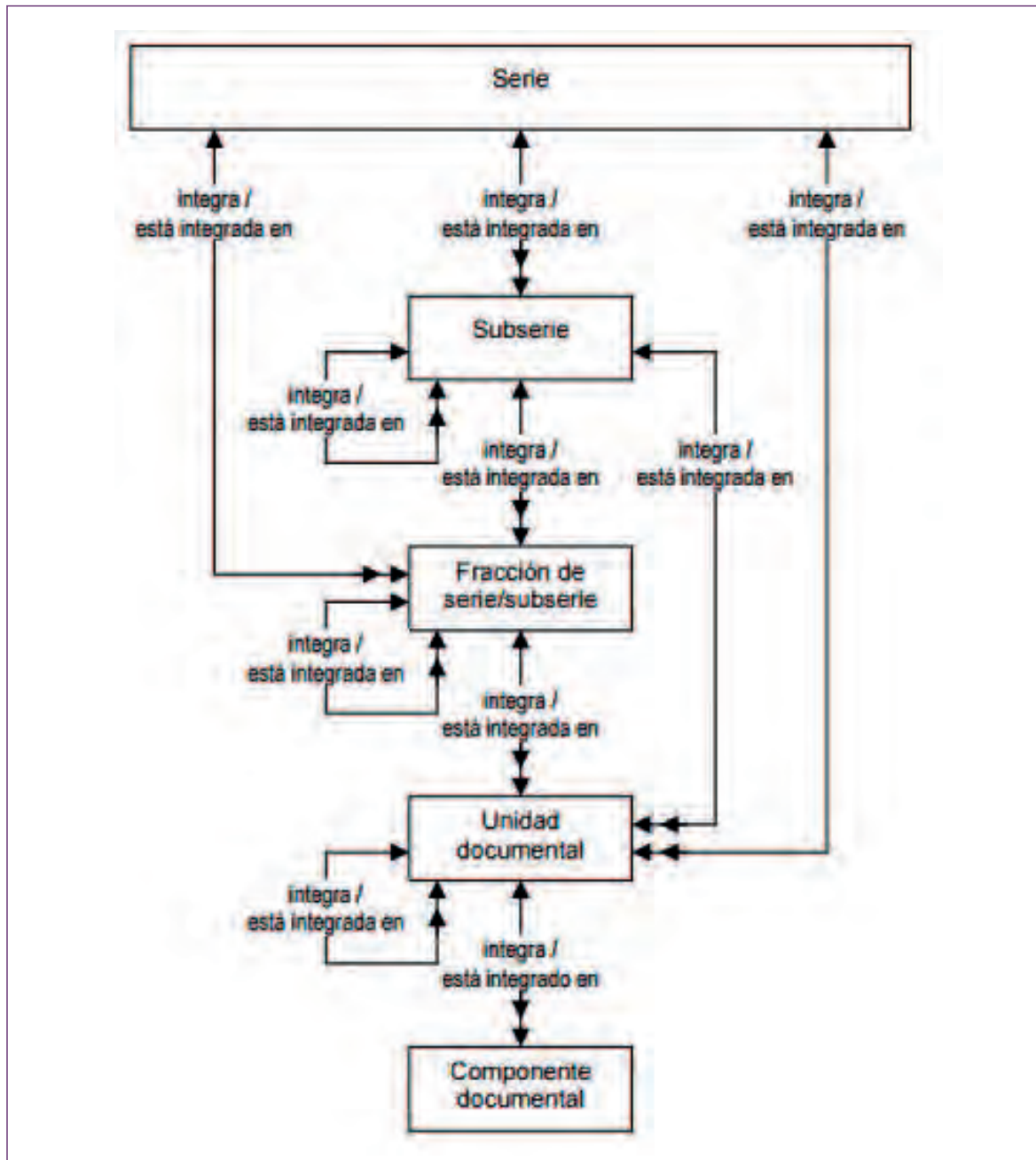


Figura3. Modelització ER (entitat-relació) de la sèrie documental en el Model Conceptual de Descripció Arxivística de les NEDA.

Finalment, l'any 2012, el CIA va constituir un grup de treball internacional, integrat per més d'un centenar de professionals de tretze països, anomenat EGAD (Experts Group on Archival Description), amb la missió de crear un model conceptual arxivístic internacional partint de les experiències prèvies ja esmentades. Els treballs s'han desenvolupat fins a la publicació, el setembre del 2016,

del primer esborrany amb el títol *Records in Context: A Conceptual Model for Archival Description*.³³ És probable —i seria recomanable— que, en el futur, el model conceptual s'integri com una extensió més de CIDOC CRM per facilitar la interoperabilitat —ara sí, de manera natural— completa entre arxius, biblioteques i museus a partir d'un marc de referència comú.

3. DELS MODELS CONCEPTUALS A LES TECNOLOGIES SEMÀNTIQUES

El disseny de models conceptuals, algun dels quals formalitzats ja com a ontologies, crea el marc necessari per a l'explicitació i la integració del coneixement. Des del punt de vista tecnològic, aquest esforç en el context del web semàntic es tradueix en la construcció d'eines amb llenguatges que permeten el processament lògic de la informació, com ara RDF i els seus dialectes, amb un grau d'expressivitat semàntica superior, com ara RDF Schema o OWL. Una iniciativa que, des de l'any 2007, està adquirint una importància cada vegada més rellevant entre les diferents estratègies del desplegament del web semàntic és *linked data* i la seva versió en obert pel que fa a l'ús i l'accés a les dades, *linked open data*.

Linked data i *linked open data* esdevenen un dels escenaris del desplegament de les tecnologies semàntiques d'integració de les GLAM.

La tecnologia LOD és una realitat consolidada i es troba al si de les polítiques de reutilització de dades governamentals.³⁴ El seu ús supera les barreres lingüístiques i culturals i fins i tot la Xina³⁵ o l'Iran, no des de posicionaments de transparència sinó de gestió de continguts, s'erigeixen en contribuïdors.

3.1 ORIGEN DE *LINKED DATA* I *LINKED OPEN DATA*

Entre els anys 2006 i 2009, Tim Berners-Lee (2006, 2009), considerat un dels creadors del web, va definir els principis de *linked data* (LD) i *linked open data* (LOD). LD és una solució tecnològica que ofereix un conjunt de bones pràctiques per a la publicació de dades i la seva interconnexió fent possible la interoperabilitat i la reutilització de la informació disponible al web (Bizer, 2009). Darrere de la premissa de reutilització es troba el concepte *open*, gestionat per mitjà de desenes de llicències d'ús.³⁶ Tecnològicament, les tecnologies LD per-

meten convertir les dades i els conjunts de dades en registres d'informació, amb l'objectiu de generar una base de dades global utilitzant les tecnologies del web, de manera que la informació sigui comprensible tant pels humans com per les màquines. L'objectiu és generar una xarxa de dades interconnectades a través d'enllaços qualificats i semàntics sobre els quals es pot integrar i generar nou coneixement. La plataforma de *linked data* és el web semàntic —tot i que alguns especialistes, considerem que encertadament, inverteixen els termes d'aquesta relació i ubiquen *linked data* com una capa més de l'estructura del web semàntic (D'Aquin, Motta, 2016)— i el fonament òptim de la seva tecnologia es basa en la utilització d'un codi únic per identificar els conceptes (tangibles o intangibles), anomenat URI (*uniform resource identifier*), que operi sobre el protocol HTTP per facilitar la recuperació i l'accés als conceptes i els recursos pels humans; un llenguatge per codificar i intercanviar informació, RDF, i un punt d'accés i interrogació de les dades anomenat Protocol and RDF Query Language (SPARQL) (Hyvönen, 2012).

L'any 2010, Berners-Lee va definir els principis de la LOD amb cinc nivells organitzats en una jerarquia de menys a més qualitat: penjar al web les dades en qualsevol format i amb una llicència oberta (1); publicar les dades de manera estructurada i en formats interpretables per una màquina (2); publicar les dades en formats no propietaris com ara CSV, XML, etc. (3); penjar les dades a través d'estàndards del W3C (World Wide Web Consortium) com ara RDF i SPARQL per vincular-les (4), i complir tots els requisits anteriors i vincular els conjunts de dades publicades (5). Un sisè nivell no formalitzat però present als principals projectes de LOD incideix en la necessitat d'utilitzar metadades per descriure els conjunts de dades també anomenats *datasets* (Ávila, 2014).

3.2 TECNOLOGIES DE LINKED OPEN DATA

RDF és el model principal de representació de la informació a LOD. És una derivació del model entitat-relació de les bases de dades basat en entitat-atribut-valor, que en RDF s'expressa com a subjecte-predicat-objecte (Van Hooiland, Verborgh, 2014, p. 43-44). Aquests tres valors lingüístics s'anomenen en RDF *triplets* i es visualitzen gràficament en forma de graf. Els triplets s'emmagatzemen en grans repositoris anomenats *triplestores*, que poden custodiar centenars o milers de milions d'aquestes estructures; són literalment repositoris, no de documents, sinó de pura semàntica. Els grafs RDF adopten la forma de dos nodes, un per al subjecte —l'element que es descriu— i l'altre per a l'objecte

—la descripció de l'element—, units per una línia o relació, el predicat —expressat mitjançant una metadada—, que aporta realment el significat (la semàntica) dels components del graf. El valor d'aquests elements —subjecte, predicat o objecte— pot ser literal o bé la referència d'un URI que identifica un recurs, de manera que un graf pot apuntar a informació de recursos diferents disponibles en línia i accessibles. La utilització d'URI en comptes de literals o paraules aporta avantatges als continguts, ja que els fa més portables i interoperables i redueixen l'ambigüitat terminològica, a més d'optimitzar els sistemes SEO (Search Engine Optimization).

Els grafs RFD es poden combinar entre si creant grafs complexos que poden arribar a ser de dimensions enormes com a manifestació del grau de vinculació entre recursos (Labra, 2011, p. 58-61) i convertir el web en una «base de dades universal», o el que és el mateix, en un web de dades. Aquest és el principi de la dissolució de les anomenades *sitges* d'informació, és a dir, la possibilitat d'alliberar la informació de bases de dades amb estructures i esquemes heterogenis i compartir la informació en un únic model de representació, RDF. No és casual, en aquest sentit, que una conseqüència de la generació massiva de grafs es trobi darrere de l'increment del nombre de noves bases de dades anomenades NoSQL,³⁷ més ràpides i menys rígides que les bases de dades relacionals.

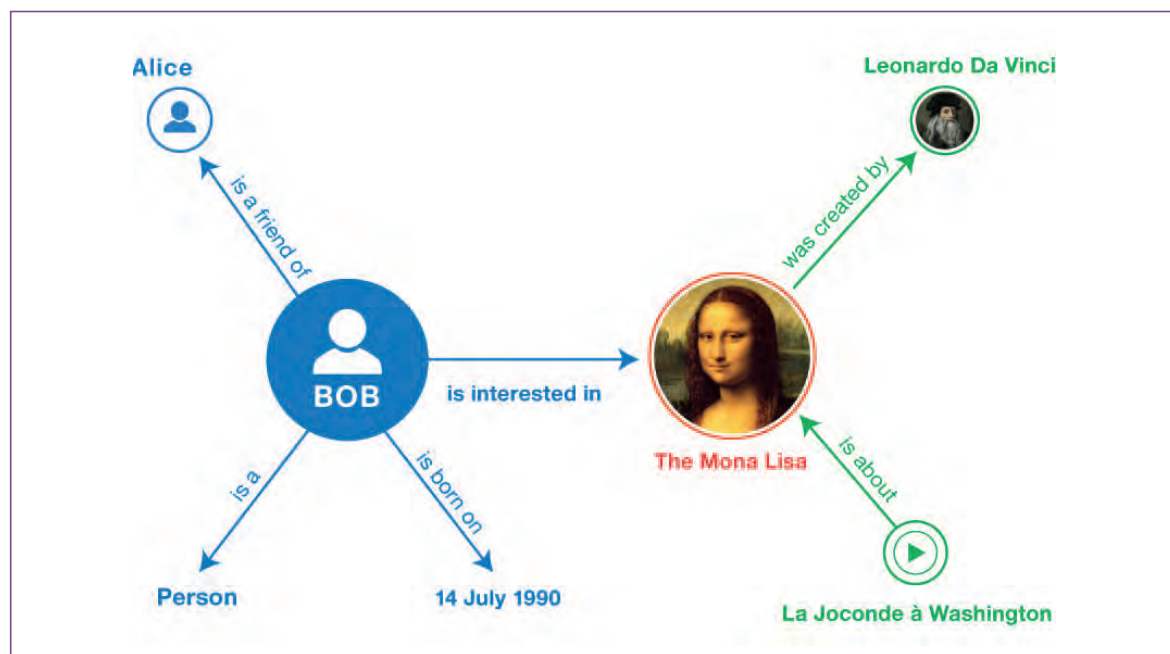


Figura 4. Conjunt de grafs —triplets RDF— que descriuen la relació entre una persona —BOB— i les seves amistats amb el quadre de la Gioconda. Les metadades provenen en la seva majoria d'esquemes externs com podria ser FOAF, del qual adopta les propietats o metadades *knows* o bé Dublin Core amb «creator» (Screiber, Raimond, 2014).

El metallenguatge RDF es pot representar de manera codificada per mitjà de la conversió dels grafs (representació gràfica d'un triplet) a diferents sintaxis (llenguatge codificat) interpretables per una màquina que reben el nom de *serialitzacions*. La més utilitzada des d'un punt de vista empresarial és RDF/XML, i Turtle és un dels formats més fàcils d'interpretar per les persones. JSON-LD és un format cridat a substituir a XML per la importància que el llenguatge JAVA està adquirint entre els desenvolupadors. Altres serialitzacions són N-Triples, TriG i RDFa; aquesta darrera serialització està adoptant un paper cada vegada més rellevant gràcies al projecte Schema.org,³⁸ fonamentat en les anomenades *microdades* o *microformats*, és a dir, dades expressades en RDF inserides a atributs d'HTML des dels quals els *parsers* o analitzadors sintàctics extreuen els triplets.

Finalment, com ja s'ha exposat prèviament, RDF és el pilar de dos llenguatges de codificació d'ontologies que expliciten i representen el coneixement: RDFS i OWL, també anomenat OWL2 per la versió actual (segona, publicada pel W3C l'any 2009).³⁹ Molts vocabularis utilitzats en el context de LOD estan expressats en una codificació o en l'altra.

3.3 MADURESA I CONSISTÈNCIA DE LES TECNOLOGIES LOD

La tecnologia que dóna suport a la iniciativa LOD gaudeix d'un nivell de desenvolupament que en algunes àrees ha assolit una maduresa molt avançada, tot i que encara no hi ha, en la majoria de casos, estàndards internacionals més enllà de les recomanacions tècniques del W3C, així com la utilització intensiva d'alguns estàndards internacionals com ara Dublin Core (ISO 15836, des del 2009) i CIDOC CRM, com ja s'ha exposat. La infraestructura de LOD es basa en part de les tecnologies del web semàntic, com també en els fonaments de l'enginyeria del programari. La irrupció de les polítiques de transparència i dades obertes a escala internacional i, també, el ràpid desenvolupament de la *big data* es troben entre els aspectes que expliquen el fort impuls que s'està experimentant darre-rament en el manteniment i la millora dels conjunts de dades publicats com a LOD en aspectes com ara la gestió del cicle de vida, el control i la garantia de la procedència, l'autenticitat, la integritat i també en el terreny de la preservació a llarg termini. Exemples no en falten. En el camp de la definició i la garantia de la qualitat dels processos que participen del cicle de vida de LOD s'han definit un seguit de metodologies,⁴⁰ entre les quals caldria destacar LOD2 (Linked Open Data Life-Cycle), un projecte cofinançat per la Comissió Europea i coordinat

per l'AKSW Research Group des de la Universitat de Leipzig.⁴¹ Aquest projecte té com a objectiu la integració de dades enllaçades a gran escala utilitzant tecnologies semàntiques per facilitar l'ús de dades de govern obert (Cabrera, 2016, p. 26) i descriu vuit grans fases, aportant en cada cas informació sobre la tecnologia, les aplicacions i els procediments aplicables (Auer, 2012). D'altra banda, des del W3C s'estan fent importants esforços per garantir la procedència, la confiança i la integritat de les dades publicades com a LOD per mitjà de la tasca portada a terme per grups com ara el W3 Provenance Working Group i el W3 WebID Incubator Group, amb la construcció d'ontologies en RDFS i OWL2 mitjançant models com ara l'Open Provenance Model (OPM)⁴² i el PROV Data Model (PROV-DM),⁴³ en el cas de la representació de la procedència; el Web Access Control Vocabulary (ACL), el Privacy Reference Ontology (PPO)⁴⁴ i el FOAF+SSL, en el camp de la seguretat i el control d'accés a les dades semàntiques, i diferents ontologies i vocabularis basats en esdeveniments per al control i l'actualització de les dades, com ara Dataset Dynamics Vocabulary (DADY), Linking Open Descriptions of Events (LODE), DSNotify Eventset Vocabulary i OWL 2 Change Ontology, entre d'altres (López *et al.*, 2011, p. 10-11).

Pel que fa a la preservació en el terreny de LOD, aportacions destacades des de l'àmbit de l'arxivística s'estan duent a terme amb el projecte Reload (Repository for Linked Open Archival Data), impulsat i mantingut per les institucions italianes Archivio Centrale dello Stato, Istituto per i Beni Artistici, Culturali e Naturali dell'Emilia-Romagna i Regesta.exe,⁴⁵ i del projecte PRELIDA (Preserving Linked Data),⁴⁶ finançat per la Comissió Europea. Ambdues aproximacions tenen per objectiu investigar diferents iniciatives internacionals com ara OAIS, Europea i DBpedia,⁴⁷ entre d'altres, per traslladar procediments, bones pràctiques i tecnologies a la preservació tant dels conjunts de dades com de la semàntica associada recollida a través de les seves vinculacions. Aquest darrer aspecte és crític, ja que s'ha de tenir present que els vincles entre *datasets* expressats com a triplets RDF es mesuren en bilions — i en constant creixement — conservats en repositoris *triplestores*, entre els quals cal destacar, per exemple, Oracle Spatial and Graph,⁴⁸ amb 1,08 bilions de triplets, i OpenLink Virtuoso v6.1, amb 15,4 bilions.

Recolzant aquests projectes hi ha una nodrida literatura científica i de difusió sobre les tecnologies al servei de LOD elaborada des del W3C,⁴⁹ el món universitari⁵⁰ i, també, des d'iniciatives de professionals del web semàntic, mitjançant, per exemple, LinkedDataTools,⁵¹ un lloc web que recull desenes i desenes d'aplicacions informàtiques i publicacions especialitzades. També cal esmentar

els centenars de tesis doctorals i treballs de fi de màster que anualment surten de facultats universitàries d'arreu del món, especialment de l'àmbit de les ciències de la computació, però cada vegada més, de l'àrea de les ciències de la informació.

3.4 **LINKED OPEN DATA A UN CLIC: APLICACIONS PRINCIPALS PER A L'ACCÉS I LA RECUPERACIÓ DE CONJUNTS DE DADES I VOCABULARIS**

El descobriment de recursos disponibles com a LOD, i la seva navegació, accés i descàrrega, es pot fer sense necessitat d'adquirir programari especialitzat a través d'un conjunt d'aplicacions desenvolupades per institucions i investigadors procedents, majoritàriament, del món acadèmic, així com del suport en matèria de recerca d'institucions públiques, com és el cas de PlanetData!,⁵² un projecte europeu de recerca per fomentar l'exposició de dades al web en formats oberts i que ha donat peu a l'elaboració de catàlegs de difusió de *datasets* en LOD, com ara el Mannheim Linked Data Catalog⁵³ i el LOD Cloud Diagram.

3.4.1 **DataHub i LOD Cloud Diagram**

El DataHub és una aplicació que permet l'accés lliure⁵⁴ al repositori *triplestore* de CKAN (Comprehensive Knowledge Archive Network),⁵⁵ una eina per a la gestió i la publicació de conjunts de dades, no necessàriament formalitzades com a LOD, utilitzada per molts governs nacionals i locals, institucions de recerca i altres organitzacions que recullen i publiquen dades obertes. Va entrar en servei l'any 2006, va ser dissenyada per l'OKF (Open Knowledge Foundation) i la versió actual és la 2.0. Actualment, el DataHub dona accés a 11.119 conjunts de dades.⁵⁶ El seu increment és notable. Així, entre el maig del 2012, amb gairebé 3.500 *datasets* (Méndez, Greenberg, 2012, p. 5), i el febrer del 2017, el seu volum s'ha més que triplicat.

L'aplicació és la cara visible de les iniciatives LOD mitjançant el popular The Linking Open Data Cloud Diagram,⁵⁷ conegut també com el núvol LOD, una representació en forma de núvol d'una part dels continguts de CKAN, els publicats com a LOD, i que compleixen uns mínims requisits de qualitat. L'actualització de la informació no és regular, sinó que els seus creadors i els responsables del manteniment l'actualitzen de manera discontinua en el temps (quatre versions l'any 2007, tres els anys 2008 i 2009, i una única versió els anys 2010,

2011, 2014 i 2017). El diagrama va acompanyat d'un informe molt interessant que actualment només està disponible per als anys 2011⁵⁸ i 2014,⁵⁹ tot i que la darrera actualització gràfica és molt recent (mitjan mes de febrer del 2017). The Cloud és explotat per altres aplicacions, com ara Linked Open Vocabularies⁶⁰ i, en l'àmbit estadístic, per LODStats,⁶¹ i el seu volum global entre la primera versió i l'actual ha passat de dotze *datasets* l'1 de maig de 2007 als 1.139 actuals.

Les aplicacions al DataHub són descrites per mitjà de metadades VOID (Vocabulary of Interlinked Datasets),⁶² un esquema en RDF específic de l'entorn LD (Keith *et al.*, 2011), que informen d'aspectes com ara les dates de creació i modificació, els responsables de la publicació, l'URI del recurs, la pàgina principal de la institució responsable del manteniment, el tipus de llicència d'ús utilitzada i els formats de presentació i descàrrega de dades, com també informació estadística sobre el nombre d'enllaços i triplets RDF, entre d'altres.

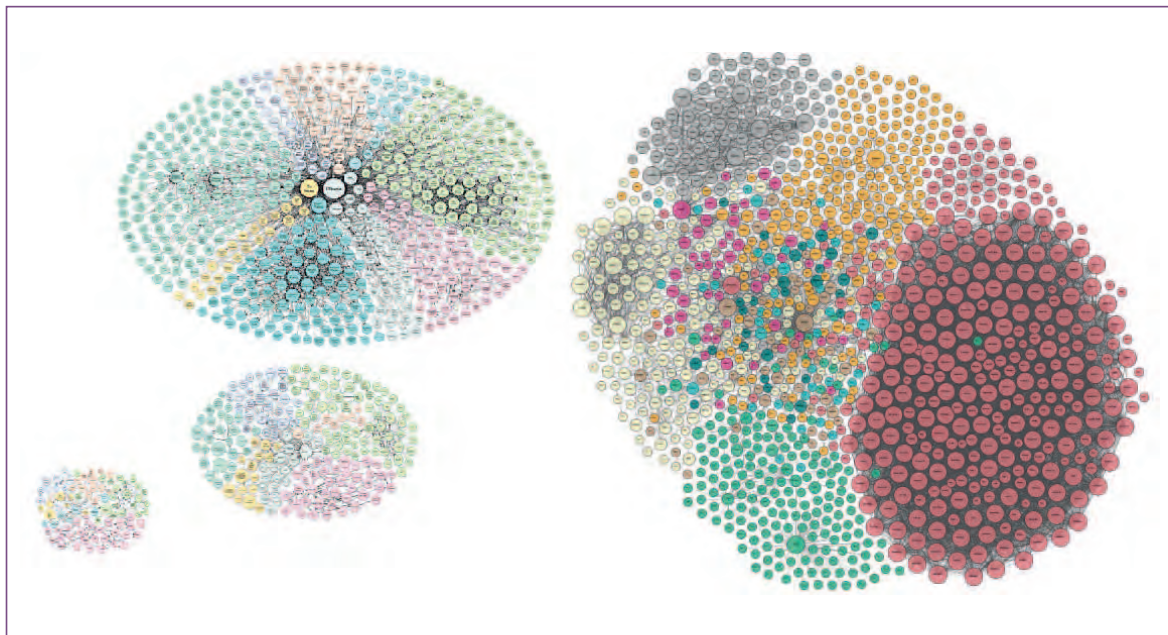


Figura 5. Evolució dels *datasets* incorporats al LOD Cloud Diagram des de l'any 2009 (95) fins al 2017 (1.146).

3.4.2 Linked Open Vocabularies (LOV)

És una aplicació que permet cercar, seleccionar i analitzar ontologies i vocabularis publicats com a LOD. Va ser creada per Bernard Vatant i Pierre-Yves Vandesbussche i publicada l'any 2012 per Mondeca Lab dins del projecte *Datalift*. Està ubicat als servidors de l'Open Knowledge Foundation (OKFN). Es tracta d'un dels projectes més innovadors i està en creixement constant. Així, el mes

de maig del 2012 registrava 212 vocabularis (Méndez, Greenberg, 2012, p. 9); menys de cinc anys després, el febrer del 2017, en dona accés a 694, tres vegades més, que sumen un total de 58.451 classes i propietats.

Organitza el contingut per tipus de vocabularis, propietats (classe, propietat, tipus de dada, instància) utilitzades en l'estructuració i la configuració dels vocabularis, com també per responsables de la seva publicació i elaboració. A més, habilita un punt d'accés i consulta SPARQL Endpoint, descàrrega a través de *data dump* en format .gz, i un catàleg d'API per accedir a vocabularis, agents i propietats.

Cada vocabulari és descrit amb metadades que aporten informació sobre el seu URI, espai de nom, la pàgina web de referència, una descripció, el llenguatge del recurs, el creador, el contribuïdor i comentaris. A més, proporciona una interfície gràfica on identifica les relacions establertes amb els vocabularis que es vinculen al recurs (Incoming Links), com també les que el recurs utilitza per a la seva configuració (Outgoing Links). Aquesta informació es completa amb dades estadístiques relatives al nombre de classes, propietats, tipus de dades i instàncies que participen del vocabulari, com també el llenguatge de representació utilitzat (Expressivity), com podria ser el cas de RDF, RDFS i OWL, i el nombre de conjunt de dades en què és utilitzat. Les classes, les propietats i les instàncies són identificades pels seus corresponents espais de nom (per exemple «owl:same – Ontology Web Language», «dcterms:title – Dublin Core» o «skos:prefLabel – SKOS Simple Knowledge Organization System Namespace Document») i identificades mitjançant el seu URI, amb el qual s'accedeix al conjunt d'elements que formen l'esquema del concepte seleccionat.

El codi font de LOV disposa d'una llicència Creative Commons Attribution 4.0 International License que permet compartir i fer modificacions indicant l'atribució, la llicència i la descripció dels canvis.

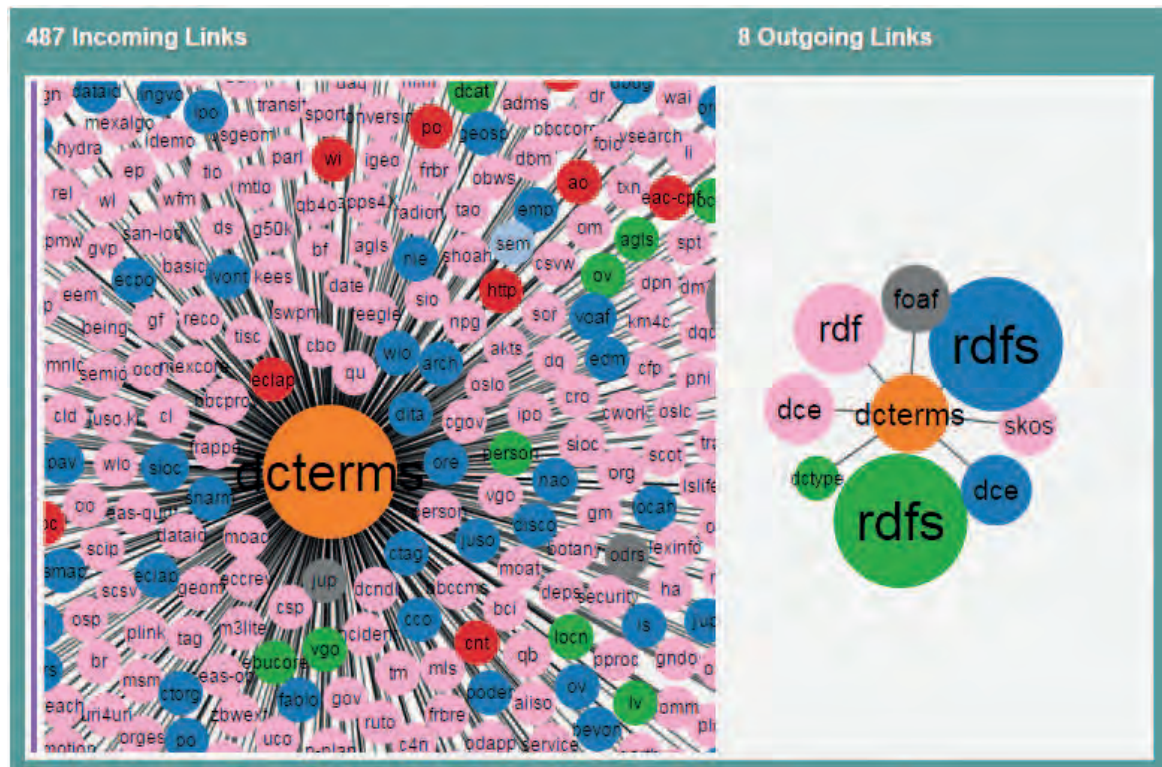


Figura 6. Visualització gràfica a través de LOV dels datasets que reutilitzen el vocabulari DCMI Metadata Terms de Dublin Core (Incoming Links), com també les ontologies i els vocabularis externs amb els quals es vincula per a la seva configuració com a publicació LOD (Outgoing Links).⁶³ Dublin Core és un dels esquemes més utilitzats i populars a LOD.

3.4.3 Open Metadata Registry (OMR)

És una de les peces fonamentals de la infraestructura tècnica del web semàntic i a hores d'ara és el registre més important per a la comunitat bibliotecària i per a altres comunitats procedents de l'àmbit científic. Inicialment es va desenvolupar per donar suport a la National Science Digital Library (NSDL),⁶⁴ tot i que en aquests moments és de lliure accés i està gestionat per la Metadata Management Associates, una associació de consultoria compromesa a mantenir el registre com un sistema obert.

El registre proporciona un mitjà per identificar, formular i publicar esquemes de metadades amb les seves propietats i perfils d'aplicació, i dona suport als procediments de descoberta de metadades, reutilització, normalització i interoperabilitat a escala global i local (Díaz, 2003, p. 20-21). Disposa d'una funcionalitat innovadora que el diferencia d'altres serveis de registre que és l'anomenat Sandbox,⁶⁵ un espai virtual de proves on qualsevol usuari pot experimentar, compartir i aprendre. Una de les seves principals creadores és la fundadora

de la comunitat Dublin Core sobre gestió de metadades (DCMI-VMC o Dublin Core Metadata Initiative – Vocabulary Management Community) (Méndez, Greenberg, 2012, p. 9).

Entre els vocabularis registrats no en consta cap que sigui propi de la comunitat arxivística (com podria ser EAD i EAC), tot i que sí que es troben disponibles bona part dels més utilitzats per les biblioteques com ara esquemes de la RDA, FRBR, MARC, UNIMARC, Dublin Core, etc.

3.5 LOD DINS DEL PATRIMONI CULTURAL

Iniciatives per incentivar la participació de les institucions europees en LOD

Europa ocupa un lloc destacat en el desplegament de tecnologies semàntiques. Són nombroses les iniciatives d'integració del patrimoni cultural que els darrers quinze anys han donat a llum projectes com ara CulturASampo,⁶⁶ DigiCULT,⁶⁷ CASPAR,⁶⁸ ResearchSpace,⁶⁹ WISSKI,⁷⁰ CLAROS⁷¹ i, especialment per la seva projecció i transcendència, Europeana, que ha incentivat la integració de recursos en LOD al si del seu model de dades EDM mitjançant projectes com ara Athena, Athena Plus i Linked Heritage, des d'on s'han dut a terme campanyes importants per a la dotació d'eines i recursos orientats a la *skosificació* de vocabularis controlats (Leroi, [2010]) o la utilització de LIDO (Lightweight Information Describing Objects),⁷² un esquema de recol·lecció de metadades desenvolupat per a l'àmbit dels museus (Tsalapati, 2012), entre moltes altres estratègies. La complexitat del tractament dels continguts de les institucions del patrimoni cultural —que conceptualment engloben totes les manifestacions de l'ésser humà al llarg de la història— les converteix en un camp d'experimentació privilegiat per la seva riquesa i transversalitat, les seves necessitats i els reptes de futur, tot això imbricat amb les enormes implicacions socials que se'n deriven (Vavliakis, 2011).

Les comunitats que formen les LAM (*libraries, archives and museums*) porten a terme un paper destacat en la publicació de conjunt de dades i vocabularis d'acord amb els principis de la LOD, especialment les biblioteques, on institucions de referència com ara la Library of Congress han transformat els seus vocabularis controlats. La necessitat de fonts de metadades expressives i altament consensuades s'ha convertit en una oportunitat per als seus professionals (Harper, Tillet, 2007, p. 49-51), que en molts casos participen des del començament en projectes com ara Dublin Core Metadata Initiative. A més, hi ha comunitats

molt dinàmiques que creixen a l'empara de les anomenades *digital humanities* amb una forta implantació i presència en el món acadèmic dels Estats Units i, en general, del món anglosaxó, on mètodes computacionals s'apliquen a la recerca en humanitats.⁷³ Un exemple n'és el LODLAM (Linked Open Data in Libraries, Archives and Museums), un col·lectiu obert a la col·laboració de professionals de les institucions de la memòria, fundat l'any 2010 per Jon Voss,⁷⁴ amb l'objectiu de compartir experiències i promoure la formació i l'ocupació professional d'arxivers, bibliotecaris i museòlegs en el terreny de la LOD i on és molt freqüent llegir ofertes laborals sol·licitant experts en enginyeria ontològica.

Fruit de la necessitat de promoure la participació de les LAM en LOD, el W3C a través del Grup Incubador de Dades Vinculades de Biblioteques elaborà, entre els anys 2010 i 2011, dos documents de referència que presenten una nodrida relació de recursos útils per a la creació i el consum de dades vinculades en el terreny, especialment de les biblioteques, però amb presència de projectes d'arxius i museus acompanyats d'unes reflexions sobre la utilitat i l'àmbit d'aplicació. Fins avui, constitueixen els treballs més exhaustius en aquest camp i es consideren una referència de consulta obligada.⁷⁵

De fet, en els seus principis *linked open data* no és aliè als fonaments tradicionals de la tasca d'arxius, biblioteques i museus. Tradicionalment, aquestes institucions són consumidores habituals de continguts transversals i han creat serveis importants de referència per al descobriment de descripcions sobre recursos compartits —*linked*—, han fet esforços destacats en el terreny de la disseminació i la divulgació del coneixement —*open*— i han creat instruments per a la gestió i l'accés als documents, objectes i informació —*data*. L'impuls a l'obertura de les dades de les administracions al si dels programes de govern obert ha situat les LAM en un lloc molt destacat, especialment en tot allò relacionat amb la reutilització de les dades del sector públic per mitjà de la Directiva 2013/37/UE⁷⁶ i la seva inclusió al dret espanyol amb la Llei 18/2015, de modificació de la Llei 37/2007⁷⁷ en què s'esmenten de manera explícita, per primer cop, els arxius, les biblioteques i els museus.

3.5.1 Oportunitats i amenaces

Molts són els avantatges de la publicació i la compartició de recursos en LOD, entre els quals destaquen:

- › La cerca federada de continguts entre arxius, biblioteques i museus.⁷⁸

- › L'obertura de fonts d'informació disponibles en sistemes i arquitectures heterogènies que permetin la constitució de fons i col·leccions virtuals, especialment important en els arxius i els museus, amb una gran dispersió de productors i sèries documentals.
- › La precisió, l'harmonització i la desambiguació de descripcions dispars sobre conceptes i objectes comuns mitjançant la utilització de propietats d'equivalència (`owl:SameAs`, `owl:equivalentClass`), dependència (`rdfs:subClassOf` o `rdfs:subPropertyOf`) o bé en el cas de les ontologies formalitzades amb SKOS – exactitud en el valor semàntic entre conceptes o classes a través de `skos:exactMatch`.
- › El foment de la interoperabilitat, la millora de la qualitat de les dades i de les seves relacions, aspecte fonamental en l'aplicació de les tecnologies de la *big data* i, per tant, de l'exploració i l'explotació computacionals de la informació.
- › L'enriquiment semàntic dels fons i les col·leccions i el descobriment de nous continguts, com també, per primera vegada, l'elaboració d'instruments d'indexació i control de vocabularis descentralitzats o bé la integració i l'alineament dels ja existents. L'elaboració d'eines, com ara classificacions i tesaurus, a partir de la reutilització d'altres de ja existents permet estalviar costos i esforços i evitar redundàncies i inconsistències, amb efectes beneficiosos sobre la reducció de programes onerosos d'harmonització semàntics. La ISO 25964⁷⁹ —la nova norma de tesaurus— sobre interoperabilitat amb altres vocabularis esdevé una eina indispensable en aquest nou enfocament distribuït, encara més tenint present que els seus fonaments faciliten la implementació i la utilització de l'ontologia SKOS per a la modelització i l'intercanvi entre sistemes de classificació i control de vocabulari (Sánchez-Cuadrado *et al.*, 2012, p. 233).

Ara bé, hi ha molts aspectes que impliquen un repte per a la qualitat i la confiança en les dades:

- › L'existència d'un gran nombre d'enllaços trencats.
- › La manca de manteniment dels recursos publicats o bé la publicació de conjunts de dades per qüestions més de visibilitat que no pas operatives.
- › La conversió a tecnologies LOD sense millorar excessivament la qualitat de les dades, com podria ser el cas de la conversió massiva i heterogènia de recursos a SKOS (Simple Knowledge Organization System) sense una revisió exhaustiva o amb poques vinculacions entre tesaurus, per exem-

ple (Pastor *et al.*, 2012, p. 251-252), o bé de la incorporació parcial de RDF a les col·leccions de les institucions patrimonials (Sulé *et al.*, 2015, p. 7-8), com també la utilització limitada d'ontologies i procediments d'alineament semàntic en la gestió dels continguts de molts repositoris digitals internacionals.⁸⁰

- › La descripció insuficient amb metadades VOID⁸¹ dels recursos publicats com a LOD.
- › L'enorme complexitat en la realització de consultes amb SPARQL Endpoint, un llenguatge de consulta sobre els grafs RDF equivalent, en la seva funció, al SQL (Structured Query Language) de les bases de dades relacionals.

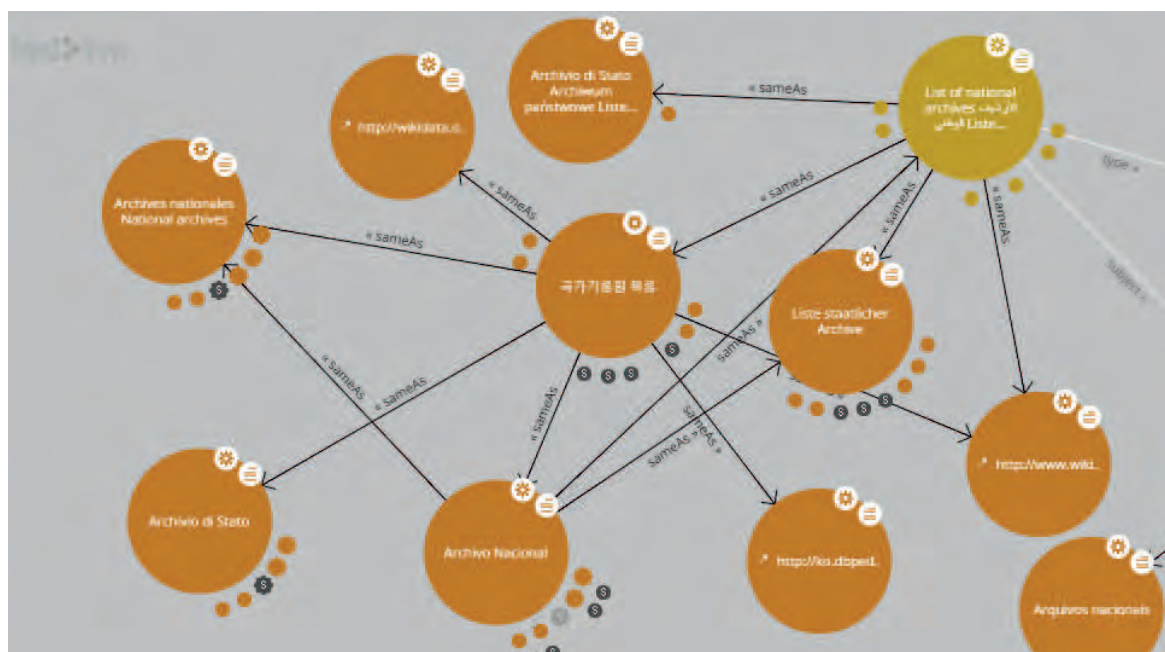


Figura 7. Concepte *arxiu nacional* de la DBpedia harmonitzat amb altres declaracions amb el mateix significat en diferents llengües fet amb vincles definits a través de la propietat owl:SameAs. Eina de visualització utilitzada: Lodlive.⁸²

Això esdevé una oportunitat indiscutible però també un repte, que obliga els professionals a un replantejament des del punt de vista de la formació, encara més quan la competència en el terreny de l'elaboració de taxonomies i instruments d'indexació per part de perfils professionals aliens als tècnics de les LAM s'ha convertit ja en una amenaça real, que podria arribar a qüestionar el seu paper i la seva necessitat, especialment dins de l'àmbit empresarial, on companyies com ara Amazon i Google desenvolupen i apliquen eines de recuperació i gestió d'informació pròpies i altament sofisticades. Els professionals de la informació han de defensar el seu paper com a garants del context⁸³ i, fins i tot, s'haurien

de transformar, amb la col·laboració d'altres perfils, en una policia semàntica que vetlli per la qualitat —i l'honestedat— de fonts d'informació i ontologies en el marc del web. Creiem que tenim moltes oportunitats que en el cas de l'arxivística passen no tan sols per publicar continguts en LOD, sinó també per fer aportacions en matèria de modelització d'ontologies que ja s'estan desenvolupant i implantant, tal com ja s'ha exposat, per a la gestió de la procedència i l'autenticitat dels recursos disponibles al web semàntic.

3.5.2 Exemples d'institucions culturals que publiquen les seves dades com a LOD

Cada vegada en són més les institucions del patrimoni cultural que exposen les seves dades en LOD i que, a hores d'ara, sumen centenars de recursos i vocabularis. En aquest esforç sobresurten especialment les biblioteques, seguides d'algunes destacades aportacions dels museus; el cas dels arxius és desolador, amb molt poques experiències, la majoria de les quals estan limitades a la publicació d'eines de control de vocabulari. A continuació se n'esmenten alguns exemples.

3.5.2.1 Europeana i la Digital Public Library of America (DPLA)

Europeana,⁸⁴ la gran plataforma d'integració del patrimoni cultural europeu, és un bon exemple de projecte orientat a la vinculació de recursos digitals amb LOD i tecnologies del web semàntic.⁸⁵ Els seus impulsors, conscients d'una banda de la necessitat de millorar la integració amb les diverses comunitats del patrimoni cultural i, de l'altra, d'incrementar la qualitat de les dades pròpies i les procedents dels agregadors en un entorn cada cop més orientat a la reutilització de la informació i l'accés obert a les dades, van decidir substituir el model de dades existent des del seu inici, Europeana Semantic Elements (ESE), per un altre model consistent en una ontologia d'alt nivell anomenada Europeana Data Model (EDM), una estructura més expressiva semànticament capaç d'integrar els diferents models de metadades sense pèrdua d'informació, mantenint en tot moment la perspectiva descriptiva de cada conjunt de dades. La primera versió es va implementar el maig del 2011 (versió 5.2.2) i, dos anys després, s'havia dut a terme la migració des d'ESE amb l'entrada en funcionament de la versió 5.2.4 (Hernández, Agenjo, 2013, p. 21-22). El nou model adopta l'enfocament orientat a esdeveniments de CIDOC CRM (Antoine, 2014, p. 17-19)⁸⁶ mitjançant la incorporació de classes contextuais com ara *agent*, *event*, *place*, *time span* i *concept*,

fet que li permet gestionar un nivell molt superior de desagregació de dades respecte d'ESE a través també de l'adopció del protocol OAI-ORE (OAI Object & Exchange) per a la representació d'objectes complexos (Doerr *et al.*, 2010, p. 5-6).⁸⁷ Amb aquest canvi, EDM pot connectar amb informació d'altres fonts de dades, possibilitar l'accés addicional sobre altres persones, llocs o conceptes i ampliar les seves funcionalitats cap a la realització de suggeriments avançats, la descoberta de nous continguts per part de l'usuari i, especialment, l'enriquiment semàntic (Juanes, 2015, p. 7). El recurs Europeana Linked Open Data està disponible al DataHub des del 28 de gener de 2015 i integra 1.806.753.903 triples RDF.⁸⁸ A més, Europeana disposa d'un seguit d'API⁸⁹ per accedir, explorar i descarregar els seus continguts i fer desenvolupaments específics.⁹⁰

D'altra banda, la Digital Public Library of America (DPLA) constitueix el projecte equivalent d'Europeana als Estats Units i adopta, amb la col·laboració d'Europeana, la tecnologia i la filosofia del model EDM (Rudersdorf, Gore, 2013, p. 3).

3.5.2.2 Lista de Encabezamientos de Materia para Bibliotecas Públicas (LEM)

És un projecte de la Subdirección General de Coordinación Bibliotecaria present com a recurs LOD el DataHub des del 20 de juliol de 2011,⁹¹ després de la conversió prèvia a l'ontologia SKOS dels seus elements, com també d'altres llistes d'encapçalaments de matèries autonòmiques com ara la catalana (LEMAC), la gallega (LEMAG) i la basca (LEMAV), amb les quals es vincula (Martínez, 2013, p. 110-114), juntament amb altres llistes internacionals com ara RAMEAU, Gemeinsame Normdatei (GND) i Library of Congress Subject Headings (LCSH).⁹² El *dataset*, constituït per 193.865 triplets RDF, disposa d'un punt d'accés SPARL Endpoint.

3.5.2.3 Library of Congress (LC)

El LC Linked Data Service - Authorities and Vocabularies ofereix desenes de vocabularis controlats i esquemes de codificació desenvolupats per aquesta institució publicats com a *linked open data* i disponibles tant des de la seva pàgina web⁹³ com des del DataHub.⁹⁴ Destaquen el referent LCSH (Library of Congress Subject Headings)⁹⁵ i l'ontologia Bibframe,⁹⁶ dissenyada a partir de l'any 2011 (Torre *et al.*, 2015, p. 115), amb la finalitat de substituir l'estàndard de catalogació Marc21.

3.5.2.4 Getty Vocabularies

Com en el cas de la LC, el Getty Research Institute ha transformat els seus tesaurs TGN (Getty Thesaurus of Geographic Names),⁹⁷ ULAN (Union List of Artist Names),⁹⁸ CONA (Cultural Objects Name Authority)⁹⁹ i ATT (Art & Architecture Thesaurus)¹⁰⁰ en LOD al llarg dels anys 2014 i 2015 amb llicència ODC-BY 1.0. Tots estan disponibles també des de la seva pàgina web.¹⁰¹

3.5.2.5 L'Encoded Archival Context (EAC-CPF) i altres projectes arxivístics

És un dels pocs vocabularis arxivístics publicats com a LOD en actiu.¹⁰² Un altre projecte, LOCAH, que partia de l'Encoded Archival Description (EAD), es va mantenir operatiu entre l'agost del 2010 i el juliol del 2011. Actualment, LOCAH, tot i que és un recurs accessible, no disposa de cap manteniment.¹⁰³ EAC-CPF en LOD ha estat desenvolupat per l'Istituto per i Beni Artistici, Culturali e Naturali en un projecte iniciat l'any 2010 i desenvolupat amb OWL, posteriorment enriquit amb altres vocabularis com ara Geonames i elements de RDF, RDFS, Dublin Core Metadata Terms i SKOS. Està format per deu classes i 29 propietats i és accessible des de Linked Open Vocabularies.¹⁰⁴

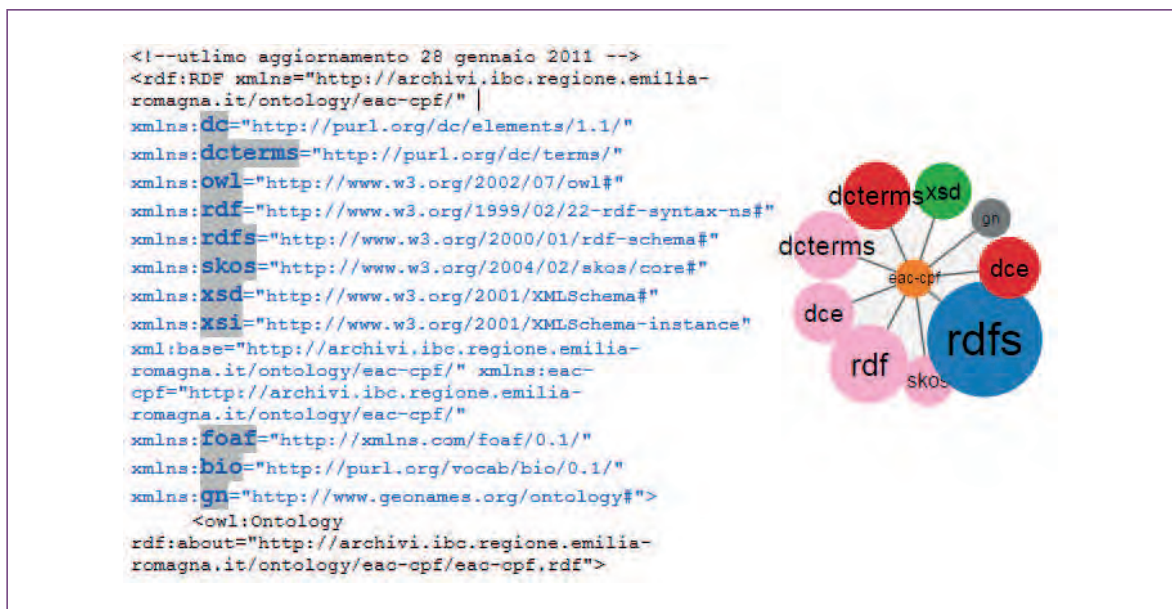


Figura 8. Superposició d'imatges corresponents a la declaració de *namespaces* o espais de noms a l'ontologia EAC-CPF (esquerra) i segment de la captura de pantalla del LOV corresponent a la mateixa ontologia on es representen gràficament (dreta).

La comunitat arxivística explora el terreny per fer noves aportacions. És un projecte que pretén transformar les dades de la descripció arxivística en *linked data* és LiAM (Linked Archival Metadata),¹⁰⁵ finançat des de l'any 2012 per l'Institute

of Museum and Library Services National Leadership Grant Program i liderat per la Tufts University.¹⁰⁶ Els seus membres proposen la necessitat d'incorporar les tecnologies semàntiques per gestionar la informació arxivística en entorns dinàmics superant les limitacions que imposa en aquest terreny l'EAD. Un dels seus productes és una interessant i completa guia per a arxivers sobre *linked data*: *Linked Archival Metadata: A Guidebook*, versió 0.99 (Lease, 2014).

3.5.2.5 Altres recursos del patrimoni cultural com a LOD

Cada dia són més les institucions culturals que difonen els seus recursos en LOD, entre les quals destaquen les biblioteques nacionals d'Espanya,¹⁰⁷ Alemanya¹⁰⁸ i Anglaterra,¹⁰⁹ o en el cas dels museus, el Rijksmuseum d'Àmsterdam¹¹⁰ i el British Museum de Londres, a través de CIDOC CRM.¹¹¹

3.5.2.6 Fonts de dades reutilitzades dins l'àmbit del patrimoni cultural

Més enllà dels tesaurs, llistats d'encapçalament de matèries o vocabularis de metadades en LOD, hi ha disponibles conjunts de dades utilitzades a bastament en procediments d'enriquiment semàntic i descobriment de recursos per part de nombroses institucions i projectes de l'àmbit de les LAM. És el cas de DBpedia,¹¹² la versió semàntica que estructura els recursos textuais de la Wikipedia utilitzada per enriquir descripcions de persones, llocs i organitzacions; Geonames,¹¹³ ontologia geogràfica que ofereix milions de topònims; Virtual International Authority File (VIAF),¹¹⁴ una font d'autoritats de persona o organització creada per l'Online Computer Library Center (OCLC)¹¹⁵ a partir de la contribució de més de 40 biblioteques nacionals, entre les quals hi ha la Biblioteca Nacional de Catalunya i la Biblioteca Nacional d'Espanya,¹¹⁶ o FOAF (Friend of a Friend),¹¹⁷ una ontologia que descriu persones, les seves activitats i les relacions amb altres persones en entorns de xarxes socials,¹¹⁸ entre molts altres *datasets*. Europea reutilitza molts d'aquests vocabularis en les seves tasques d'enriquiment i manteniment de la qualitat de les dades.¹¹⁹

4. CONCLUSIONS A TALL DE REFLEXIÓ DES DE L'ANÀLISI DE L'APLICACIÓ DE LOD ALS ARXIUS

Tal com s'ha exposat, entre les comunitats de les LAM, l'arxivística és la més introvertida en la publicació de conjunts de dades com a LOD. Es podrien avan-

çar quatre raons. En primer lloc, els arxius estan limitats extremadament per dependències organitzatives fins i tot dins d'una mateixa administració, cosa que dificulta enormement la col·laboració en l'elaboració d'instruments comuns i, també, en la seva reutilització. En segon lloc, a diferència de l'àmbit de les biblioteques, amb més de 140 anys d'experiència en l'elaboració d'instruments d'indexació (la classificació decimal de Dewey data del 1876), el recorregut de les eines de descripció internacionals arxivístiques no ha superat el quart de segle. En tercer lloc, els arxivers som un col·lectiu amb reptes que haurien de ser abordats des de plantejaments propis d'una enginyeria però que ens formem en àrees allunyades dels laboratoris universitaris, en especial dels de les facultats de computació. Finalment, les comunitats del patrimoni cultural i, en general, les que integrem les ciències de la informació, malauradament, coexistim des del desconeixement generalitzat tant de les oportunitats que ofereixen la integració dels nostres recursos en la gestió diària com en la defensa dels nostres interessos professionals comuns.

Aprofundint en el tercer punt, l'arxivística constitueix una disciplina particular en el conjunt de les LAM, ja que és l'única que aplica i integra criteris d'enginyeria —potser no formalment, però sí des de la pràctica— en els seus processos quotidians (anàlisi de circuits, disseny i gestió de quadres de classificació, etc.). Aquesta singularitat implica que la seva presència al web no ha de ser exclusivament cultural sinó també de serveis, és a dir, com a proveïdors d'eines estratègiques per al funcionament de les administracions, tan ben formalitzades que els seus components haurien d'estar preparats per a una reutilització constant. Creiem que aquest és el motiu principal de l'absència dels arxius a LOD. Els arxivers, malauradament, actuem en moltes ocasions —i ens hi proclamem oficialment— com a gestors de documents i realment no ho som —no pas de manera exclusiva—, en absolut; som arquitectes, analistes de sistemes i no exclusivament de documents sinó, i especialment, d'informació. La iniciativa LOD i la nostra responsabilitat en una administració i societat cada vegada més obertes i transparents ens porta a una bifurcació: cedir el pas a altres perfils professionals i replegar-nos exclusivament, per honestedat, a la funció tradicional dels arxius —rellevant i necessària, d'altra banda— amb la seguretat que ofereix l'enfocament docucèntric o bé reformular el nostre sistema de formació —parlem des d'una perspectiva general i molt enfocada a les *digital humanities* anglosaxones— des del qual reivindicar-nos professionalment com a perfils necessaris i honestament compromesos amb les competències que hem de desenvolupar.¹²⁰ No podem predicar mediàticament la necessitat del nostre col·lectiu en la garantia de la transparència sense constituir-nos en actius tecnolò-

gics indispensables en la seva aplicació més enllà de la gestió documental, que certament és una peça fonamental. Altrament, estem creant falses expectatives.

Al nostre entendre, tres serien els àmbits d'aplicació de LOD als arxius més enllà de puntuals vocabularis controlats o instruments de descripció.¹²¹ Principalment i molt necessaris serien els esquemes de metadades, cada vegada més abundants i nodrits d'informació davant la implantació de l'administració electrònica. Com en el cas dels quadres de classificació —segon instrument necessari com a LOD—, l'atomització en la seva generació és excessiva, encara més si tenim present que, pel que fa a l'administració local, amb competències molt ben definides, per posar un exemple, cada institució elabora les seves pròpies eines. Això comença a generar la necessitat d'harmonitzar, d'establir equivalències, d'organitzar, des del respecte a la diversitat, la constel·lació de recursos disponibles amb criteris d'interoperabilitat i de reaprofitament. En tercer lloc, les taules d'avaluació documental i els instruments d'accés, transversals en moltes administracions en la seva aplicabilitat i execució, també s'haurien de publicar en aquesta tecnologia. Però no és qüestió de transformar vells instruments en tecnologies innovadores, sinó de repensar la forma de treballar. La majoria d'institucions creen i gestionen els seus recursos de manera no estructurada. Els arxivers som professionals de les estructures que utilitzem, paradoxalment, eines desestructurades en el nostre dia a dia —hi tornem a insistir, per l'orientació formativa i el pes de la tradició historicista. No generem productes orientats a la seva reutilització. La publicació com a LOD d'aquests instruments basats en procediments d'enginyeria tant en la confecció com en el manteniment —en processos necessàriament iteratius— reduiria els enormes esforços de generació paral·lela de productes i permetria corregir moltes de les deficiències que imposen les barreres organitzatives i competencials.

Relacionat amb aquesta perspectiva, els quadres de classificació —en tant que eina troncal, estratègica i sistèmica— s'haurien de dissenyar en capes d'informació altament estructurada cercant l'aprofitament de fonts de dades externes per integrar elements o enriquir semànticament els existents o, fins i tot, per crear nous quadres de classificació a partir de l'aprofitament d'elements d'altres esquemes publicats com a LOD, d'acord amb els principis de la norma ISO 25964 i, fins i tot també, per establir les bases d'una hibridació amb els fonaments conceptuals d'altres col·lectius professionals.¹²²

Afortunadament, ja existeixen projectes orientats a la integració i la reutilització d'eines, sobretot en aquelles institucions que de manera transversal apleguen

competències en matèria d'arxius, biblioteques i museus, com seria el cas del Museu Marítim de Barcelona, que ha desenvolupat un projecte ambiciós i robust anomenat GIM, una eina que, segons la nostra opinió, és fàcilment escalable a LOD (García-Morales, 2011).¹²³ D'altra banda, recentment s'ha presentat la primera versió d'un quadre de classificació per a la documentació municipal impulsat des de la Subdirecció General d'Arxius i Museus i dissenyat amb un enfocament altament estructurat i orientat a la reutilització dels seus components.¹²²

NOTES

- 1 En aquesta secció es presenten els models conceptuals d'abast més general tot i que existeixen d'altres amb objectius més específics i concrets. Un exemple especialment rellevant, en aquest sentit, és el model harmonitzat per a la producció de dades culturals (Modèle harmonisé pour la production des données culturelles) orientat a la integració i interoperabilitat d'informació procedent del patrimoni cultural francès. El model va ser elaborat en el marc del programa HADOC (Harmonisation de la production des données culturelles) a càrrec del Ministeri de Cultural i Comunicació francès l'any 2008. L'ontologia està formalitzada en UML - Unified Modeling Language- i és compatible amb altres ontologies com CIDOC CRM i EMD així com esquemes de metadades com LIDO (Briatte, 2013). Actualment és un dels instruments de referència a França per a l'elaboració de programes d'exposició de dades com a Open Linked Data en el context del web semàntic.
- 2 ISO 21127:2006. Information and documentation – A reference ontology for the interchange of cultural heritage information. Disponible a: <https://www.iso.org/standard/34424.html> [Consulta: 18 de desembre de 2016].
- 3 ISO 21127:2014. Information and documentation – A reference ontology for the interchange of cultural heritage information. Disponible a: <https://www.iso.org/standard/57832.html> [Consulta: 18 de desembre de 2016].
- 4 Definition of the CIDOC Conceptual Reference Model. Version 6.2.2. January 2017. Disponible a: http://www.cidoc-crm.org/sites/default/files/2017-01-25%23CIDOC%20CRM_v6.2.2_esIP.pdf. [Consulta: 18 de desembre de 2016].
- 5 <http://protege.stanford.edu/products.php> [Consulta: 18 de desembre de 2016].
- 6 <http://erlangen-crm.org/> [Consulta: 18 de desembre de 2016].
- 7 <http://lov.okfn.org/dataset/lov/vocabs/event> [Consulta: 10 de gener de 2017].
- 8 DSNotify Eventsets: A vocabulary for change events in linked data sources (dsn). Disponible a: <http://lov.okfn.org/dataset/lov/vocabs/dsn> [Consulta: 26 de gener de 2017].
- 9 LODE (Linking Open Descriptions of Events). Disponible a: <http://lov.okfn.org/dataset/lov/vocabs/lode> [Consulta: 26 de gener de 2017].
- 10 <http://lov.okfn.org/dataset/lov/vocabs/dul>.
- 11 <http://www.opencyc.org/> [Consulta: 10 de gener de 2017].
- 12 Disponible a la pàgina web de CIDOC CRM http://LOD.cidoc-crm.org/mapping_technology.html [Consulta: 27 de gener de 2017]. L'esquema XSD de la proposta es pot consultar a http://139.91.183.3/mapping_technology/xsd/x3ml/x3ml_v1.0.xsd [Consulta: 27 de gener de 2017].
- 13 Tant la pàgina web de CIDOC com les de moltes altres institucions acadèmiques ofereixen

múltiples exemples de mapatges amb altres esquemes de metadades i models de dades; en el cas de CIDOC CRM, es pot consultar a: http://LOD.cidoc-crm.org/crm_mappings.html [Consulta: 15 de desembre de 2016]. Alguns exemples concrets de mapatges amb CIDOC CRM són els relatius a LIDO (Lightweight Information Describing Objects) a través de casos descriptius com el publicat per Koutraki-Doerr (2010); altres exemples són Arachne Project, UNIMARC, a través de Dublin Core, MIDAS Standard, Dublin Core Element Set, model de dades AMICO, EAD (Bountouri, Gergatsoulis, 2011), VRA Core 4.0 o el model de dades Europeana Data Model. En altres casos, CIDOC CRM s'ha utilitzat com una plataforma d'integració simultània de diferents esquemes de metadades procedents dels tres grans àmbits del patrimoni cultural, a fi i efecte de promoure una gestió integrada dels seus recursos: EAD —arxius—, MARC —biblioteques— i LIDO —museus— (Bargudo, 2013) o, finalment, en el context del projecte europeu SCULPTEUR (Le Boeuf, 2005).

- 14 Són molts els projectes que utilitzen l'extensió CRMdif com a base del raonament semàntic en la gestió, per exemple, de repositoris (Tzompanaki, Doerr, 2012 i [2013]) o bé en la creació d'ontologies destinades a facilitar la gestió de demandes d'informació per part d'usuaris dels arxius en llenguatge natural en el context del model AKM (Archival Knowledge Model) (Hennicke, 2013).
- 15 <http://www.cidoc-crm.org/crminf/> [Consulta: 10 de gener de 2017].
- 16 <http://www.cidoc-crm.org/crmarchaeo/> [Consulta: 10 de gener de 2017].
- 17 <http://www.cidoc-crm.org/crmsci/> [Consulta: 10 de gener de 2017].
- 18 <http://www.cidoc-crm.org/crmgeo/> [Consulta: 10 de gener de 2017].
- 19 <http://www.cidoc-crm.org/crmdig/> [Consulta: 10 de gener de 2017].
- 20 <http://www.cidoc-crm.org/crmba/> [Consulta: 10 de gener de 2017].
- 21 <http://www.cidoc-crm.org/frbroo/> [Consulta: 10 de gener de 2017].
- 22 <http://www.cidoc-crm.org/presso/> [Consulta: 10 de gener de 2017].
- 23 Disponible en versió castellana a <http://www.ifla.org/files/assets/cataloguing/frbr/frbr-es.pdf> [Consulta: 10 de febrer de 2017] i en català a http://www.ifla.org/files/assets/cataloguing/frbr/frbr-ca_2009.pdf [Consulta: 10 de febrer de 2017].
- 24 Existeix la traducció al castellà realitzada per la Comisión de Traducción de la Biblioteca Nacional de España: http://www.ifla.org/files/assets/cataloguing/frad/frad_2009-es.pdf [Consulta: 10 de febrer de 2017]. La versió catalana, traduïda per Assumpció Estivill Rius, es va publicar l'any 2010: http://www.ifla.org/files/assets/cataloguing/frad/frad_2009-ca.pdf [Consulta: 10 de febrer de 2017].
- 25 La versió de revisió es pot descarregar a http://www.ifla.org/files/assets/cataloguing/frbr-lrm/frbr-lrm_20160225.pdf [Consulta: 1 de març de 2017] i va acompanyada d'un altre document anomenat Transition Mappings, una eina de suport a l'avaluació del nou model que estableix correspondències entre cada element dels models conceptuals unificats. http://www.ifla.org/files/assets/cataloguing/frbr-lrm/transitionmapping_20160225.pdf [Consulta: 1 de març de 2017].
- 26 Library of Congress. Bibframe. Disponible a: <https://www.loc.gov/bibframe/> [Consulta: 10 de febrer de 2017].
- 27 Una anàlisi molt completa i actualitzada de les darreres iniciatives experimentals en el camp de l'aplicació d'eines semàntiques a través de *linked data* en el camp de les biblioteques sobre la base de BIBFRAME i, especialment, FRBR s'exposa a Godby et al. (2015). OCLC és l'acrònim de The Online Computer Library Center, una organització sense ànim de lucre orientada a la recerca en el terreny de l'accés informatitzat a recursos bibliogràfics.
- 28 <http://visualdataweb.de/webvowl/> [Consulta: 17 de febrer de 2017].
- 29 Finnish Conceptual Model for Archival Description, draft version 01: <http://www.arkisto.fi/uploads/Arkistolaitos/Hankkeet/AHAA/The%20Finnish%20Conceptual%20Model%20of%20Archi->

- [val%20Description_text.pdf](#) [Consulta: 20 de gener de 2017].
- 30 PROYECTO NEDA-I (Normas Españolas de Descripción Archivística – Proyecto I: Normas de estructura y contenido de datos para las descripciones de documentos de archivo, agentes, funciones y archivos), http://www.mecd.gob.es/cultura-mecd/dms/mecd/cultura-mecd/areas-cultura/archivos/mc/cneda/recursos/documentos-apoyo/20070917_Proyecto-NE-DAI/20070917_Proyecto%20NEDA_I.pdf [Consulta: 19 de gener de 2017].
 - 31 Modelo Conceptual de Descripción Archivística y requisitos de datos básicos de las descripciones de documentos de archivo, agentes y funciones. Parte 1: Tipos de entidad. Parte 2: Relaciones, http://www.mecd.gob.es/dms/mecd/cultura-mecd/areas-cultura/archivos/mc/cneda/documentacion/NEDA_MCDA_P1_P2_20120618-pdf/NEDA_MCDA_P1_P2_20120618.pdf [Consulta: 19 de gener de 2017].
 - 32 Modelo Conceptual de Descripción Archivística. Parte 3: Atributos, [http://www.mecd.gob.es/dms/mecd/cultura-mecd/areas-cultura/archivos/mc/cneda/documentacion/MC-DA-P3--15-01-2016-_borradorIP/MCDA-P3%20\(15-01-2016\)_borradorIP.pdf](http://www.mecd.gob.es/dms/mecd/cultura-mecd/areas-cultura/archivos/mc/cneda/documentacion/MC-DA-P3--15-01-2016-_borradorIP/MCDA-P3%20(15-01-2016)_borradorIP.pdf) [Consulta: 19 de gener de 2017].
 - 33 Records in Context: A Conceptual Model for Archival Description. Consultation Draft v0.1. September 2016: <http://www.ica.org/sites/default/files/RiC-CM-0.1.pdf> [Consulta: 19 de gener de 2017].
 - 34 Així, a Espanya, per exemple, el web oficial de dades obertes del Govern espanyol hi destina una secció completa (<http://datos.gob.es/es/blog-tags/linked-data> [Consulta: 19 de gener de 2017]), al mateix temps que el portal homònim de la Generalitat de Catalunya en fa esment (http://dadesobertes.gencat.cat/ca/que_es_l_open_data/concepte/ [Consulta: 10 de gener de 2017]). Al Regne Unit, el portal *data.gov.uk* ofereix 41.183 datasets, molts dels quals com a LOD (https://data.upf.edu/en/linked_data [Consulta: 10 de gener de 2017]). Arreu del món, i especialment a Europa, els projectes de foment d'utilització de LOD formen part d'un ecosistema d'iniciatives que abasta amplis sectors no exclusivament polítics, sinó també, i especialment, acadèmics, sanitaris, culturals, empresarials i socials. La Universitat Pompeu Fabra (https://data.upf.edu/en/linked_data [Consulta: 10 de gener de 2017]), per exemple, forma part del projecte Linked Universities (<http://linkeduniversities.org/> [Consulta: 10 de gener de 2017]), que fomenta la publicació i la reutilització del coneixement entre universitats europees a partir de LOD.
 - 35 El projecte *Zhishi.me* (Niu et al., 2011), des de l'any 2011, s'obre a la disposició d'ingents quantitats de dades com a LOD per mitjà del descobriment automatitzat d'enllaços en recursos massius com enciclopèdies en línia.
 - 36 Una aproximació bastant completa a l'ampli espectre de les llicències d'ús i reutilització de recursos utilitzades en el context d'open data es pot consultar a Korn (2009). D'altra banda, Europeaana utilitza catorze tipus de llicències (<http://pro.europeana.eu/share-your-data/rights-statement-guidelines/available-rights-statements> [Consulta: 17 de febrer de 2017]). Una altra font indispensable per a l'anàlisi pràctica de la utilització de llicències és la consulta de les metadades VOID o informació de context dels conjunts de dades publicats al DataHub, un repositori de recursos open data, bona part dels quals són publicats com a LOD (<https://datahub.io/dataset> [Consulta: 28 de gener de 2017]).
 - 37 NoSQL és la definició d'un tipus de sistemes de gestió de bases de dades no relacionals que operen al marge del llenguatge de consulta SQL (*structured query language*). La informació deixa d'agrupar-se en taules i esquemes preconcebuts per ser organitzada, entre altres procediments, en forma de grafs. Hi ha un gran nombre d'aplicacions que les utilitzen destinades, fonamentalment, a la gestió d'enormes volums de dades com ara Facebook, Twitter i LinkedIn o buscadors com ara Google i Yahoo (Joyanes, 2014, p. 185-189).
 - 38 *Schema.org* (<http://schema.org/>) és una iniciativa fruit de la col·laboració dels buscadors web Google, Bing i Yahoo per a la millora de l'etiquetatge, la interpretació del contingut i la millora del posicionament de les pàgines web. L'any 2013, RDFa representà el segon microformat més important després d'*HTML-mf-hcalendar*, amb un total de 471.406 dominis i més de 2.600 milions de triples (<http://webdatacommons.org/structureddata/2013-11/stats/stats.html>)

- [Consulta: 6 de març de 2017].
- 39 <https://www.w3.org/TR/2009/REC-owl2-overview-20091027/> [Consulta: 6 de març de 2017].
- 40 Una completa visió del desplegament de metodologies per a la gestió integral del cicle de vida de les dades governamentals publicades com a LD — Publishing Government Linked Data— es recull a Villazón et al. (2011).
- 41 <http://lod2.eu/Welcome.html> [Consulta: 6 de març de 2017].
- 42 <http://lov.okfn.org/dataset/lov/vocabs/opmv>.
- 43 El model PROV-DM s'ha representat com a ontologia a partir d'OWL2 i ha donat com a resultat la PROV-O (Lebo et al., 2013).
- 44 <http://lov.okfn.org/dataset/lov/vocabs/ppo> [Consulta: 1 de març de 2017].
- 45 <http://labs.regesta.com/progettoReload/en/> [Consulta: 1 de març de 2017].
- 46 <http://pro.europeana.eu/blogpost/spotlight-on-prelida-preserving-linked-data> [Consulta: 28 de desembre de 2016].
- 47 PRELIDA. Preserving Linked Data. ICT-2011. 4.3: Digital Preservation. D3.1 State of the art assessment on Linked Data and Digital Preservation [en línia] <<http://www.prelida.eu/sites/default/files/D3.1%20State%20of%20the%20art.pdf> [Consulta: 18 de febrer de 2017].
- 48 Oracle Spatial and Graph: Benchmarking a Trillion Edges RDF Graph. Oracle white paper. November 2016. Disponible a: http://download.oracle.com/otndocs/tech/semantic_web/pdf/OracleSpatialGraph_RDFgraph_1_trillion_Benchmark.pdf [Consulta: 1 de març de 2017].
- 49 El W3C posa a disposició dels desenvolupadors un ampli catàleg de recomanacions tècniques, relacions d'aplicacions informàtiques així com manuals de bones pràctiques, entre els quals, en el terreny del LOD, destaca *Best Practices for Publishing Linked Data*, de Hyland et al. (2014).
- 50 Són moltes les publicacions tècniques sorgides de l'àmbit universitari, entre les quals destaca el llibre *Linked Data: Envolving the Web into a Global Data Space*, publicat per Morgan i Claypool l'any 2011 i que s'ha convertit en tot un referent, entre altres coses pel fet de ser una de les primeres publicacions especialitzades que oferien una visió transversal de la iniciativa LOD. Un extens resum es troba en línia a Heath et al. (2011).
- 51 LinkedDataTools.com [Consulta: 1 de març de 2017].
- 52 <http://www.planet-data.eu/mission> [Consulta: 1 de març de 2017].
- 53 <http://linkeddatacatalog.dws.informatik.uni-mannheim.de/ca/>.
- 54 <https://datahub.csail.mit.edu/www/> [Consulta: 1 de març de 2017].
- 55 <https://ckan.org/> [Consulta: 1 de març de 2017].
- 56 <https://datahub.io/es/dataset> [Consulta: 1 de març de 2017].
- 57 La versió actual *Linked Open Data cloud diagram 2017* ha estat creada per Abele et al. (2017).
- 58 http://lod-cloud.net/state/state_2011/ [Consulta: 1 de març de 2017].
- 59 http://lod-cloud.net/state/state_2014/ [Consulta: 2 de gener de 2017]. Els informes han servit de base per a la realització d'anàlisis estadístiques per part de diferents comunitats de les GLAM. Un exemple des de l'àmbit bibliotecari és el treball de Ríos (2015, p. 35-47).
- 60 <http://lov.okfn.org/dataset/lov> [Consulta: 1 de març de 2017].
- 61 <http://stats.lod2.eu/> [Consulta: 1 de març de 2017].
- 62 <http://lov.okfn.org/dataset/lov/vocabs/void> [Consulta: 2 de gener de 2017].
- 63 <http://lov.okfn.org/dataset/lov/vocabs/dcterms> [Consulta: 20 de febrer de 2017].

- 64 <https://nsoi.oercommons.org/> [Consulta: 2 de gener de 2017].
- 65 <http://sandbox.metadatarregistry.org/> [Consulta: 1 de març de 2017].
- 66 <http://www.kulttuurisampo.fi/?lang=en> [Consulta: 17 de febrer de 2017].
- 67 <http://www.digicult.info/pages/info.php> [Consulta: 17 de febrer de 2017].
- 68 <http://www.dcc.ac.uk/resources/briefing-papers/technology-watch-papers/caspar>.
- 69 <http://www.researchspace.org/>.
- 70 <http://wiss-ki.eu/>.
- 71 <https://www.beazley.ox.ac.uk/PID1023719.pdf>.
- 72 <http://network.icom.museum/cidoc/working-groups/lido/what-is-lido/>.
- 73 Un enllaç molt descriptiu per entendre l'abast dels objectius de la formació en *digital humanities* es pot trobar a la pàgina web de la Universitat d'UCLA, a Los Angeles, Estats Units (<http://www.cdh.ucla.edu/about/contact/> [Consulta: 1 de març de 2017]) o, amb més detall, a Drucker et al. (2014).
- 74 <http://lodlam.net/about/> [Consulta: 1 de desembre de 2016].
- 75 Els documents són: *Informe final del Grupo Incubador de Datos Vinculados de Bibliotecas* (Baker et al., 2011), disponible en castellà gràcies a la traducció de l'anglès de Xavier Agenjo i Francisca Hernández, i *Conjuntos de datos vinculados, vocabularios de valores y conjuntos de elementos de metadatos* (Isaac et al., 2011), traducció de l'anglès al castellà a càrrec de Juan Antonio Pastor i Tomás Saorín.
- 76 https://www.boe.es/diario_boe/txt.php?id=DOUE-L-2013-81251 [Consulta: 5 de gener de 2017].
- 77 <https://www.boe.es/boe/dias/2015/07/10/pdfs/BOE-A-2015-7731.pdf> [Consulta: 5 de gener de 2017].
- 78 Un bon exemple de projecte de cerca federada sobre arxius, biblioteques i museus basada en l'ontologia de CIDOC el podem trobar a Hernández (2008).
- 79 ISO 25964-1:2011 Information and documentation -- Thesauri and interoperability with other vocabularies -- Part 1: Thesauri for information retrieval (<https://www.iso.org/standard/53657.html>) i ISO 25964-2:2013 Information and documentation -- Thesauri and interoperability with other vocabularies -- Part 2: Interoperability with other vocabularies (<https://www.iso.org/standard/53658.html>) [Consulta: 1 de març de 2017].
- 80 Això es posa especialment de manifest als informes de la COAR (Confederation of Open Access Repositories). Un exemple és l'informe del 21 de maig de 2015, *Promoting Open Knowledge and Open Science Report of the Current State of Repositories* (COAR, 2015, p. 12).
- 81 Un gran nombre dels recursos publicats a través del DataHub no disposen de metadades VOID o en tenen una descripció molt defectuosa.
- 82 <http://en.lodlive.it/> [Consulta: 4 de març de 2017].
- 83 Citant Stefan Gradmann (2012), Méndez ([2014]) afirma que «[...] las bibliotecas deben pasar de ser contenedoras a contextualizadoras de la información».
- 84 <http://www.europeana.eu/portal/ca> [Consulta: 25 de febrer de 2017].
- 85 <http://labs.europeana.eu/api/linked-open-data-introduction> [Consulta: 25 de febrer de 2017].
- 86 Es tracta de la versió castellana traduïda per Domingo Arroyo Fernández de l'original anglès *Europeana Data Model Primer*.
- 87 A partir de la definició i la gestió dels *proxies* procedents d'OAI-ORE en Europeana, s'han fet conversions d'EAD a EDM per la complexitat que representa la jerarquia en què es representen els objectes arxivístics, tal com exposen Hennie et al. (2011, p. 85-88).

- 88 <https://datahub.io/dataset/europeana-lod> [Consulta: 15 de desembre de 2016].
- 89 <http://labs.europeana.eu/api> [Consulta: 25 de febrer de 2017]
- 90 Accessible des de la pàgina web d'Europeana: <http://labs.europeana.eu/> [Consulta: 15 de desembre de 2016].
- 91 <https://datahub.io/dataset/lista-encabezamientos-materia> [Consulta: 15 de desembre de 2016].
- 92 <http://id.sgcb.mcu.es/> [Consulta: 15 de desembre de 2016].
- 93 <http://id.loc.gov/> [Consulta: 15 de desembre de 2016].
- 94 <https://datahub.io/dataset/lcsh> [Consulta: 15 de desembre de 2016].
- 95 <https://datahub.io/dataset/lcsh> [Consulta: 15 de desembre de 2016].
- 96 <http://lov.okfn.org/dataset/lov/vocabs?q=bibframe> [Consulta: 15 de desembre de 2016].
- 97 <https://datahub.io/dataset/getty-tgn> [Consulta: 15 de desembre de 2016].
- 98 <https://datahub.io/dataset/getty-ulan> [Consulta: 15 de desembre de 2016].
- 99 <https://datahub.io/dataset/getty-cona> [Consulta: 15 de desembre de 2016].
- 100 <https://datahub.io/dataset/getty-aat> [Consulta: 15 de desembre de 2016].
- 101 <http://www.getty.edu/research/tools/vocabularies/lov/> [Consulta: 15 de desembre de 2016].
- 102 <http://archivi.ibc.regione.emilia-romagna.it/ontology/semanticEnvironment/index.html> [Consulta: 15 de desembre de 2016].
- 103 <http://data.archiveshub.ac.uk/about.html> [Consulta: 15 de desembre de 2016].
- 104 <http://lov.okfn.org/dataset/lov/vocabs/eac-cpf> [Consulta: 15 de desembre de 2016].
- 105 <http://sites.tufts.edu/liam/background/> [Consulta: 15 de desembre de 2016].
- 106 <https://tramullas.com/liam-linked-archival-metadata-para-archiveros/> [Consulta: 15 de desembre de 2016].
- 107 La transformació a LOD es va dur a terme amb la implicació de l'Ontology Engineering Group de la Universitat Politècnica de Madrid, una de les institucions científiques més prestigioses del món en el terreny del desenvolupament d'eines semàntiques. Per a la seva consulta, vegeu el recurs que descriu el procés de transformació a Vila et al. (2012).
- 108 http://www.dnb.de/EN/Service/DigitaleDienste/LinkedData/linkedata_node.html.
- 109 Es pot consultar el procés de transformació a *linked open data* d'una part de la col·lecció de la British National Library a través de la publicació de Corine (2014). D'altra banda, des del Mannheim Linked Catalog, però també des de la pàgina de dades obertes del Govern britànic, l'accés a la informació com a LOD es troba disponible. Des de la pàgina DATA.GOV.UK: <https://data.gov.uk/dataset/the-linked-open-british-national-bibliography> [Consulta: 7 de març de 2017]; des del Mannheim Linked Catalog: <http://linkeddatacatalog.dws.informatik.uni-mannheim.de/dataset/bluk-bnb-basic> [Consulta: 7 de març de 2017].
- 110 <http://www.semantic-web-journal.net/content/rijksmuseum-collection-linked-data-2>.
- 111 <http://collection.britishmuseum.org/>.
- 112 <https://datahub.io/dataset/dbpedia> [Consulta: 15 de desembre de 2016].
- 113 <https://datahub.io/dataset/geonames> [Consulta: 15 de desembre de 2016].
- 113 <https://datahub.io/dataset/viaf> [Consulta: 15 de desembre de 2016].
- 114 <http://www.oclc.org/en/home.html> [Consulta: 15 de desembre de 2016].
- 115 <https://viaf.org/> [Consulta: 15 de desembre de 2016].

- 116 <http://lov.okfn.org/dataset/lov/vocabs/foaf> [Consulta: 15 de desembre de 2016].
- 117 <http://www.foaf-project.org/> [Consulta: 15 de desembre de 2016].
- 118 Un document de referència sobre els processos i les tecnologies d'enriquiment semàntic a Europea es pot consultar a *Report on Enrichment and Evaluation*, publicat el 10 d'octubre de 2015 i disponible en línia a: http://pro.europeana.eu/files/Europeana_Professional/EuropeanaTech/EuropeanaTech_taskforces/Enrichment_Evaluation/FinalReport_EnrichmentEvaluation_102015.pdf [Consulta: 15 de desembre de 2016].
- 119 L'apel·lació a la incorporació d'una formació amb una forta base computacional no és pròpia d'alguns sectors del col·lectiu arxivístic. En altres àmbits com el dels documentalistes i els bibliotecaris, més preparats que el nostre col·lectiu des del punt de vista de les competències informàtiques, es fan formulacions semblants. Un cas paradigmàtic és el d'Isidro Aguillo, coordinador del Laboratorio de Cibermetría del Centro de Ciencias Sociales y Humanas (CCHS) del CSIC, que defensa la introducció de la lingüística computacional en els currículums dels màsters de l'àmbit de les facultats de biblioteconomia i documentació (Aguillo, 2012, p. 7).
- 120 En essència, un quadre de classificació és una eina sistèmica que integra funcions, descripcions, legislació, taules d'avaluació, unitats orgàniques, perfils professionals, nivells d'accés i seguretat, referències temàtiques, temporals i espacials relatius a fets —esdeveniments que conformen el context—, recursos d'informació relacionats o eines de control de vocabulari com ara tesaurus per facilitar la recuperació de la informació des de múltiples perspectives. Bona part de les fonts d'informació que podrien proveir aquestes capes de sucosa semàntica ja es troba disponible en forma de LOD. A través del cercador del DataHub es pot accedir a centenars de classificacions realitzades amb finalitats principalment estadístiques, econòmiques i socials per part d'institucions governamentals d'àmbit nacional però també internacional, que proveeixen de conceptes que defineixen funcions i que poden ser perfectament reutilitzats. A més, hi ha altres recursos com ara Geonames per a referències geogràfiques, DBpedia per a informació descriptiva, BIBO (The Bibliographic Ontology) per a vinculacions amb recursos bibliogràfics, ontologies legals en diferents àmbits com ara el German Labor Thesaurus, que està vinculat amb el tesaurus Eurovoc i, per tant, hereta un enfocament multilingüe, etc. D'altra banda, la conversió a SKOS de la taxonomia de la classificació i les seves relacions obre la porta a reaprofitar o enllaçar altres conceptes d'altres classificacions, no necessàriament arxivístiques. El desplegament d'ontologies en llenguatges semàntics, especialment en OWL, que modelitzin i recullin la informació sobre l'accés i l'avaluació de la documentació permetria no tan sols la seva vinculació als quadres de classificació en LOD, sinó també l'aplicació de tècniques d'automatització i execució d'operacions altament complexes i transversals en moltes administracions, gràcies al disseny de regles que serien processades per raonadors semàntics. Quadres de classificació i instruments relacionats formalitzats en OWL obririen la porta a una gestió documental eficient i segura on bona part dels processos massius que actualment es duen a terme mecànicament es podrien automatitzar i on la consistència del mateix sistema de gestió documental estaria garantida per regles d'inferència. Propagada aquesta perspectiva a la totalitat del cicle de vida de la informació, la integració amb plataformes com Europeana seria un procés més i executat de manera natural i, no cal dir-ho, el servei a les administracions des de l'enfocament de la transparència seria realment una garantia.
- 121 Cada vegada són més nombroses les transformacions de tesaus i instruments de control de vocabulari a les tecnologies LD. Un cas paradigmàtic és, dins de l'àmbit de l'administració local francesa, la conversió a OLD del tesauere per a l'indexació d'arxius locals (Sibile, Motte, 2014, p. 1-11).
- 121 Tot i que el projecte es remunta al 2010, la presentació de l'eina d'integració de la informació procedent de les col·leccions de l'arxiu, la biblioteca i el museu del Museu Marítim de Barcelona es presentà durant la 6a Jornada de Museus, Gestió Documental i Arxiu el març del 2016.
- 123 Departament de Cultura de la Generalitat de Catalunya. Publicació del Quadre de Classificació per a la documentació municipal (QdCM) http://cultura.gencat.cat/ca/detall/Noticies/Noticia_QdCM_definitiu [Consulta: 20 de setembre de 2016].

BIBLIOGRAFIA

- › ABELE, Andreis; MACCRAE, John P.; BUITELAAR, Paul; JENTZSCH, Anja; CYGANIAK, Richard. *Linking Open Data cloud diagram 2017* [en línia]. 2017. <<http://lod-cloud.net/>> [Consulta: 1 de març de 2017].
- › AGENJO, Xavier; HERNÁNDEZ, Francisca. «Los modelos de datos en la perspectiva de Europea, la DPLA y RDA» [en línia]. *Boletín ANABAD*. Núm. 3 (2013). <https://www.academia.edu/13252179/Los_modelos_de_datos_en_la_perspectiva_de_Europeana_la_DPLA_y_RDA> [Consulta: 15 de desembre de 2016].
- › AGUILLO, Isidro F. «Tecnologías, investigación y futuro de la profesión». *El Personal de la Información*. Vol. 21 (gener-febrer del 2012). <http://www.elprofesionaldelainformacion.com/contenidos/2012/enero/index.html> [Consulta: 21 de gener de 2017].
- › ARKISTOLAITOS ARKIVVERKET. *Finnish Conceptual Model For Archival Description, draft version 01* [en línia]. <http://www.arkisto.fi/uploads/Arkistolaitos/Hankkeet/AHAA/The%20Finnish%20Conceptual%20Model%20of%20Archival%20Description_text.pdf> [Consulta: 20 de gener de 2017].
- › AUER, Sören; BÜHMANN, Lorenz; DIRSCHL, Christian; ERLING, Orri; HAUSENBLAS, Michael; ISELE, Robert; LEHMANN, Jens; MARTIN, Michael; MENDES, Pablo N.; VAN NUFFELEN, Bert; STADLER, Claus; TRAMP, Sebastian; WILLIAMS, Hugh. «Managing the Life-Cycle of Linked Data with the LOD2 Stack» [en línia]. Dins: *The Semantic Web – ISWC 2012*. Berlín: CU-DRÉ-MAUROUX, P. et al. (eds.), 2012. P. 1-16 <<http://iswc2012.semanticweb.org/sites/default/files/76500001.pdf>> [Consulta: 7 de març de 2017].
- › ÁVILA ALONSO, Rafael. *Aplicación de los principios Linked Open Data a la lista de encabezamientos de materia de la Biblioteca de la Universidad Politécnica de Madrid* [en línia]. Treball de màster. Madrid: Universidad Carlos III de Madrid, 2014. 256 p. <http://orff.uc3m.es/bitstream/handle/10016/19674/avila_aplicacion_TFM_2014.pdf?sequence=1> [Consulta: 15 de gener de 2017].
- › BARBUDO CARRASCO, Lais. «Information Integration: Mapping Cultural Heritage Metadata into CIDOC CRM» [en línia]. Dins: *XXV Congresso Brasileiro de Biblioteconomia, Documentação e Ciência da Informação*, 7-10 de juliol de 2013. Florianópolis, Brasil. <<https://portal.febab.org.br/anais/article/viewFile/1409/1410>> [Consulta: 15 de desembre de 2016].
- › BAKER, Thomas; BERMÈS, Emmanuelle; COYLE, Karen; DUNSIRE, Gordon; ISAAC, Antoine; MURRAY, Peter; PANZER, Michael; SCHNEIDER, Jodi; SINGER, Ross; SUMMERS, Ed; WAITES, William; YOUNG, Jeff; ZENG, Marcia. *Informe final del Grupo Incubador de Datos Vinculados de Bibliotecas. Informe del Grupo Incubador del W3C de 25 de octubre de 2011* [en línia]. <<http://www.larramendi.es/LAM/Incubador/ld/XGR-ld-20111025.html>> [Consulta: 1 de febrer de 2017].
- › BERNERS-LEE, Tim. *Linked Data* [en línia]. 2006. <<https://www.w3.org/DesignIssues/LinkedData.html>> [Consulta: 10 de gener de 2017].
- › BERNERS-LEE, Tim. «The next Web of open, linked data» [en línia]. *TED: Ideas Worth Spreading* (febrer del 2009). <https://www.ted.com/talks/tim_berniers_lee_on_the_next_web> [Consulta: 6 de març de 2017].
- › BIZER, Christian. *Linked Data - The history so far* [en línia]. Special Issue on Linked Data, International Journal on Semantic Web and Information Systems (IJSWIS), 2009. 26 p. <<http://tomheath.com/papers/bizer-heath-berniers-lee-ijswis-linked-data.pdf>> [Consulta: 6 de març de 2017].
- › BOUNTOURI, Lina; GERGATSOULIS, Manolis. «Mapping Encoded Archival Description to CIDOC CRM» [en línia]. Dins: *First Workshop on Digital Information Management*, 2001. <<https://lekythos.library.ucy.ac.cy/bitstream/handle/10797/13028/02.Bountouri.pdf?sequence=1&isAllowed=y>> [Consulta: 15 de desembre de 2016].
- › CABRERA LOAYZA, María del Carmen. *Conceptualización y generación de RDF de los recursos bibliográficos de la Biblioteca Benjamín Carrión de UTPL. Tesis del Máster Universitario en Inteligencia Artificial de la Escuela Técnica Superior de Ingenieros Informáticos* [en línia]. Madrid:

- Universidad Politécnica de Madrid, 2016. P. 26. <<http://oa.upm.es/40329/>> [Consulta: 6 de març de 2017].
- › COMISIÓN DE NORMAS ESPAÑOLAS DE DESCRIPCIÓN ARCHIVÍSTICA. *Proyecto NEDA-I (Normas Españolas de Descripción Archivística – Proyecto I: Normas de estructura y contenido de datos para las descripciones de documentos de archivo, agentes, funciones y archivos)* [en línia]. 2007. <http://www.mecd.gob.es/cultura-mecd/dms/mecd/cultura-mecd/areas-cultura/archivos/mc/cneda/recursos/documentos-apoyo/20070917_Proyecto-NEDAI/20070917_Proyecto%20NEDAI.pdf> [Consulta: 19 de gener de 2017].
 - › COMISIÓN DE NORMAS ESPAÑOLAS DE DESCRIPCIÓN ARCHIVÍSTICA. *Modelo Conceptual de Descripción Archivística y requisitos de datos básicos de las descripciones de documentos de archivo, agentes y funciones. Parte 1: Tipos de entidad. Parte 2: Relaciones* [en línia]. 2012. <http://www.mecd.gob.es/dms/mecd/cultura-mecd/areas-cultura/archivos/mc/cneda/documentacion/NEDA_MCDA_P1_P2_20120618-pdf/NEDA_MCDA_P1_P2_20120618.pdf> [Consulta: 19 de gener de 2017].
 - › COMISIÓN DE NORMAS ESPAÑOLAS DE DESCRIPCIÓN ARCHIVÍSTICA. *Modelo Conceptual de Descripción Archivística. Parte 3: Atributos* [en línia]. 2016. <[http://www.mecd.gob.es/dms/mecd/cultura-mecd/areas-cultura/archivos/mc/cneda/documentacion/MC-DA-P3--15-01-2016-_borradorIP/MCDA-P3%20\(15-01-2016\)_borradorIP.pdf](http://www.mecd.gob.es/dms/mecd/cultura-mecd/areas-cultura/archivos/mc/cneda/documentacion/MC-DA-P3--15-01-2016-_borradorIP/MCDA-P3%20(15-01-2016)_borradorIP.pdf)> [Consulta: 19 de gener de 2017].
 - › D'AQUIN, Mathieu; MOTTA, Enrico. *The Epistemology of Intelligent Semantic Web Systems. Synthesis Lectures on The Semantic Web: Theory and Technology*, lecture 14. Morgan&Claypool Publishers, 2016. P. 6.
 - › DELIOT, Corine. «Publishing the British National Bibliography as Linked Open Data» [en línia]. British Library Board, (2014). <http://www.bl.uk/bibliographic/pdfs/publishing_bnb_as_lod.pdf> [Consulta: 28 de febrer de 2017].
 - › DIJSHOORN, Chris; JONGMA, Lizzy; AROYO, Lora; VAN OSSENBRUGGEN, Jacco; SCHREIBER, Guus. «The Rijksmuseum Collection as Linked Data» [en línia]. *Semantic Web Journal* (by IOS Press). 2016. <<http://www.semantic-web-journal.net/system/files/swj1353.pdf>> [Consulta: 23 de gener de 2017].
 - › DÍAZ ORTUÑO, Pedro Manuel. «Problemática y tendencias en la arquitectura de metadatos web» [en línia]. *Anales de Documentación*. Núm. 6 (2003). <<http://revistas.um.es/analesdoc/article/view/2051>> [Consulta: 3 de gener de 2017].
 - › DOERR, Martin; GRADMANN, Stefan; HENNICKE, Steffen; ISAAC, Antoine; MEGHINI, Carlo; VAN DE SOMPEL, Herbert. «El Modelo de Datos de Europeana (EDM)» [en línia]. *Dins: Meeting: 149. Information Technology, Cataloguing, Classification and Indexing with Knowledge Management World Library and Information Congress. 76th Ifla General Conference and Assembly*, 10-15 d'agost de 2010. Gotemburg. <<http://www.ifla.org/past-wlic/2010/149-doerr-es.pdf>> [Consulta: 23 de novembre de 2016].
 - › DRUCKER, Johanna; KIM, David; SALEHIAN, Iman; BUSHONG, Anthony. *Introduction to Digital Humanities. Course Book. Concepts, Methods, and Tutorials for Students and Instructors* [en línia]. Los Ángeles: UCLA, 2014. <http://dh101.humanities.ucla.edu/wp-content/uploads/2014/09/IntroductionToDigitalHumanities_Textbook.pdf> [Consulta: 1 de març de 2017].
 - › GARCÍA-MORALES HUIDOBRO, Elisa. *El proyecto GIM del Museu Marítim de Barcelona: un camino hacia la gobernanza de la información*. 2011 <file:///C:/Users/ACAU/Downloads/20120503MMB_GIM-Gobernanza.pdf> [Consulta: 12 de desembre de 2016].
 - › GODBY, Carol Jean; WANG, Shenghui; MIXTER, K. Jeffrey. *Library Linked Data in the Cloud. OCLC's Experiments with New Models of Resource Description. Synthesis Lectures on The Semantic Web: Theory and Technology*. Morgan & Claypool, 2015. 154 p.

- › HARPER, Corey A.; TILLET, Barbara B. *Library Of Congress Controlled Vocabularies and Their Application to the Semantic Web* [en línia]. The Haworth Press, Inc., 2007 <http://polaris.gseis.ucla.edu/gleazer/462_readings/harper.pdf> [Consulta: 29 de novembre de 2016].
- › HARPRING, Patricia. *The Getty Vocabularies and Open Linked Data: Introduction and Editorial Perspective* [en línia]. <https://www.getty.edu/research/tools/vocabularies/Linked_Data_Getty_Vocabularies.pdf> [Consulta: 10 de gener de 2017].
- › HEATH, Tom; BIZER, Christian. *Linked Data. Envolving the Web into a Global Data Space* [en línia]. 2011. <<http://linkeddatatobook.com/editions/1.0/#htoc30>> [Consulta: 12 de desembre de 2016].
- › HENNICKE, Steffen. «Representation of Archival User Needs using CIDOC-CRM» [en línia]. Dins: *CEMEX 2013. Practical Experiences with CIDOC CRM and its Extensions co-located with the 17th International Conference on Theory and Practice of Digital Libraries (TPDL 2013)*. Valetta, Malta, 26 de setembre de 2013. <<http://ceur-ws.org/Vol-1117/paper5.pdf>> [Consulta: 11 de desembre de 2016].
- › HENNICKE, Steffen; OLESKY, Marlies; DE BOER, Victor; ISAAC, Antoine; WIELEMARKER, Jan. «Conversion of EAD into EDM Linked Data» [en línia]. *Proceedings of the 1st International Workshop on Semantic Digital Archives*, 2011. <<http://ceur-ws.org/Vol-801/paper7.pdf>> [Consulta: 23 de novembre de 2016].
- › HERNÁNDEZ CARRASCAL, Francisca. «Ontología del Patrimonio de Cantabria» [en línia]. *El Profesional de la Información*. Núm. 17 (gener-febrer del 2008). <<http://www.elprofesionaldelainformacion.com/contenidos/2008/enero/11.pdf>> [Consulta: 30 de novembre de 2016].
- › HYLAND, Bernadette; ATEMEZING, Ghislain; VILLAZÓN TERRAZAS, Boris. *Best Practices for Publishing Linked Data. W3C Working Group Note 09 January 2014* [en línia]. <<http://www.w3.org/TR/2014/NOTE-ld-bp-20140109/>> [Consulta: 27 de novembre de 2016].
- › HYVÖNEN, Eero (2012). *Publishing and Using Cultural Heritage Linked Data on the Semantic Web*. Synthesis Lectures on The Semantic Web: Theory and Technology. James Hendler and Ying Ding, Series Editors. Lecture 3.
- › IFLA. *Requisitos funcionales de los registros bibliográficos. Informe final*. Traducció de Xavier Agenjo i María Luisa Martínez-Conde [en línia]. Madrid: Ministeri de Cultura. Secretaria General Tècnica, 2004. 239 p. <<http://www.ifla.org/files/assets/cataloguing/frbr/frbr-es.pdf>> [Consulta: 10 de febrer de 2017].
- › IFLA. *Requisits funcionals de les dades d'autoritat. Un model conceptual*. Working Group on Functional Requirements and Numbering of Authority Records (FRANAR). Informe final. Traducció d'Assumpció Estivill Rius [en línia]. Barcelona: Biblioteca Nacional de Catalunya, 2010. <http://www.ifla.org/files/assets/cataloguing/frad/frad_2009-ca.pdf> [Consulta: 10 de febrer de 2017].
- › IFLA. *FRBR-Library Reference Model* [en línia]. 2016-02-21 Draft for World-Wide Review, 2016. <http://www.ifla.org/files/assets/cataloguing/frbr-lrm/frbr-lrm_20160225.pdf> [Consulta: 1 de març de 2017].
- › IFLA. *Transition Mappings User Tasks, Entities, Attributes, and Relationships in FRBR, FRAD, and FRASAD mapped to their equivalents in the FRBR-Library Reference Model*. Draft to accompany the World-Wide Review of FRBR-LRM 2016-02-22 [en línia], 2016. <http://www.ifla.org/files/assets/cataloguing/frbr-lrm/transitionmapping_20160225.pdf> [Consulta: 1 de març de 2017].
- › INTERNATIONAL COUNCIL ON ARCHIVES. *Records in Context: A Conceptual Model for Archival Description. Consultation Draft v0.1* [en línia]. September 2016. <<http://www.ica.org/sites/default/files/RiC-CM-0.1.pdf>> [Consulta: 19 de gener de 2017].
- › ISAAC, Antoine. *Manual básico de Europeana Data Model. 14/07/2013* [en línia]. <http://travesia.mcu.es/portalanb/jspui/bitstream/10421/7261/1/EDM_Primer_130714_es.pdf> [Consulta: 26 de desembre de 2016].
- › ISAAC, Antoine; WAITES, William; YOUNG, Jeff; ZENG, Marcia. *Grupo Incubador de Datos Vinculados de Biblioteca: Conjuntos de datos, vocabularios de valores y conjuntos de elementos de*

- metadatos. Informe del Grupo Incubador del W3C de 25 de octubre de 2011* [en línia]. <http://www.larramendi.es/LAM/Incubator/llid/XGR-llid-vocabdataset.html> [Consulta: 1 de desembre de 2016].
- › JOYANES AGUILAR, Luis. *Big Data. Análisis de grandes volúmenes de datos en organizaciones*. Mèxic: Alfaomega, 2014. P. 185-189.
 - › JUANES HERNÁNDEZ, César. *Enriquecimiento semántico de colecciones digitales a través del Europeaana Data Model* [en línia]. 2015. <http://www.digibis.com/images/PDF/Cesar_Juanes_Hernandez_Jornadas_AAB.pdf> [Consulta: 15 de desembre de 2016].
 - › KEITH, Alexander; CYGANIAK, Richard; HAUSENBLAS, Michael; ZHAO, Jun. *W3C. Describing Linked Datasets with the Void Vocabulary. W3C Interest Group Note 03, March 2011* [en línia]. <<https://www.w3.org/TR/void/>> [Consulta: 2 de gener de 2017].
 - › KORN, Naomi; OPPENHEIM, Charles. *Licensing Open Data: A Practical Guide. version 2.0*. [en línia]. 2011. 8 p. <http://discovery.ac.uk/files/pdf/Licensing_Open_Data_A_Practical_Guide.pdf> [Consulta: 15 de novembre de 2016].
 - › KOUTRAKI, Maria, DOERR, Martin: *Mapping LIDO v0.7 to CIDOC-CRM v5.0.1* [en línia]. *Working paper*, ICS-FORTH, (març del 2010). P. 1-12. <www.cidoc-crm.org/sites/default/files/Mapping_lido_v2.doc> [Consulta: 15 de desembre de 2016].
 - › LABRA GAYO, José Emilio. *Web Semántica. Comprendiendo el cambio hacia la Web 3.0*. Oleiros: Netbiblo, 2011. P. 58-61.
 - › LAGOZE, Carl; HUNTER, Jane. *The ABC Ontology and Model* [en línia]. National Institute of Informatics, NII, Tòquio, 24-26 (octubre del 2001). P. 160-176. <<http://dcpapers.dublincore.org/pubs/article/viewFile/655/651>> [Consulta: 26 de gener de 2017].
 - › LE BOEUF, Patrick. «Transforming FRBR into FRBRoo» [en línia]. *Ciclo Biblioteche, libri, documenti: dall'informazione alla conoscenza*, 10 de maig de 2013. Roma: Universitat La Sapienza, 2013. P. 1-14 <<http://LOD.cidoc-crm.org/docs/Transforming%20FRBR%20into%20FRBRoo.pdf>> [Consulta: 26 de gener de 2017].
 - › LE BOEUF, Patrick; SINCLAIR, Patrick; MARTÍNEZ, Kirk; LEWIS, Paul; AITKEN, Geneviève; LAHANIER, Christian. «Using an ontology for interoperability and browsing of museum, library and archive information» [en línia]. Dins: *ICOM Committee for Conservation, 14th triennial meeting, the Hague, 12-16 September 2005*. Londres: James & James. P. 1-6. <https://www.researchgate.net/publication/39995744_Using_an_ontology_for_interoperability_and_browsing_of_museum_library_and_archive_information> [Consulta: 15 de desembre de 2016].
 - › LEBO, Timothy; SAHOO, Satya; McGuinness, Deborah. *PROV Ontology W3C Recommendation 30 april 2013* [en línia]. <<https://www.w3.org/TR/prov-o/>> [Consulta: 1 de març de 2017].
 - › LEROI, Marie-Véronique. *Votre terminologie comme un élément web sémantique. Recommandations de conception et de gestion* [en línia]. Athena WP4 i Linked Heritage WP3, [2010].
 - › LÓPEZ DE IPIÑA, Diego; AGUILERA, Unai; EMALDI, Mikel; LÁZARO, Jon; BUJÁN, David; LAGO, Ana Belén; ALONSO, Ainhoa; ABAITUA, Joseba. *Estado del arte en confianza y calidad de fuentes de datos enlazadas. Proyecto BIDEI* [en línia]. Bilbao: BizkaiLab, 2011. <<https://www.bizkai-lab.deusto.es/wp-content/uploads/2012/04/5761.pdf>> [Consulta: 1 de març de 2017].
 - › MARTÍNEZ CONDE, María Luisa. «La conversión a SKOS de la Lista de Encabezamientos de Materia para Bibliotecas Públicas» [CD-ROM]. *Boletín ANABAD*. Núm. 3 (2013).
 - › MÉNDEZ, Eva; GREENBERG, Jane. «Datos enlazados para vocabularios abiertos y marco general de HIVE» [en línia]. *El Profesional de la Información* (maig-juny del 2012). P. 5. <<http://recyt.fecyt.es/index.php/EPI/article/view/33368/17917>> [Consulta: 9 de desembre de 2016].
 - › MÉNDEZ RODRÍGUEZ, Eva; BUENO DE LA FUENTE, Gema: *En el Reino de LOD LAM... Linked Open Data y Patrimonio Digital* [en línia]. [2014]. <http://e-archivo.uc3m.es/bitstream/handle/10016/22653/MendezBueno2015_es.pdf?sequence=1> [Consulta: 1 de març de 2017].

- › NIU, Xing; SUN, Xinruo; WANG, Haofen; RONG, Shu; QI, Guilin; YUL, Yong. «Zhishi.me – Weaving Chinese Linking Open Data» [en línia]. Dins: *International Semantic Web Conference. ISWC 2011: The Semantic Web – ISWC 2011*. P. 205-220. <https://link.springer.com/chapter/10.1007/978-3-642-25093-4_14> [Consulta: 7 de març de 2017].
- › ORACLE. *Oracle Spatial and Graph: Benchmarking a Trillion Edges RDF Graph. Oracle white paper. November 2016* [en línia]. <http://download.oracle.com/otndocs/tech/semantic_web/pdf/OracleSpatialGraph_RDFgraph_1_trillion_Benchmark.pdf> [Consulta: 1 de març de 2017].
- › PASTOR SÁNCHEZ, Juan Antonio; MARTÍNEZ MÉNDEZ, Francisco Javier; RODRÍGUEZ MUÑOZ, José Vicente. «Aplicación de SKOS para la interoperabilidad de vocabularios controlados en el entorno de Linked Data» [en línia]. *El Profesional de la Información*. Núm. 21 (maig-juny del 2012). <<https://recyt.fecyt.es/index.php/EPI/article/viewFile/epi.2012.may.04/17918>> [Consulta: 1 de desembre de 2016].
- › PICCO, Paola; ORTIZ REPISO, Virginia. «RDA, el nuevo código de catalogación: cambios y desafíos para su aplicación» [en línia]. *Revista Española de Documentación Científica*. Núm. 35, 1 (gener-març del 2012). P. 145-173.
<http://www.ugr.es/~rruiz/docencia/automatizada/RDA_Nuevo_codigo_de_catalogacion.pdf> [Consulta: 10 de gener de 2017].
- › PRELIDA. *Preserving Linked Data. ICT-2011. 4.3: Digital Preservation. D3.1 Satate of the art assessment on Linked Data and Digital Preservation* [en línia]. <<http://www.prelida.eu/sites/default/files/D3.1%20State%20of%20the%20art.pdf>> [Consulta: 18 de febrer de 2017].
- › RÍOS HILARIO, Ana B.; GIL URDICIAIN, Blanca. «Los datos bibliotecarios en la nube de datos: análisis de los datasets GLAM presentes en el LOD cloud diagram» [en línia]. *Scire*. Núm. 21 (juliol-desembre del 2015). P. 35-47. <www.iversid.eu/ojs/index.php/scire/article/download/.../3811> [Consulta: 2 de gener de 2017].
- › RUDERSDORF, Amy; GORE, Emily. «La colaboración multiplica el valor: el modelo de asociación internacional de la DPLA y Europea» [CD-ROM]. *Boletín ANABAD*. Núm. 3 (2013).
- › SÁNCHEZ CUADRADO, Sonia; COLMENERO RUIZ, María Jesús; MOREIRO, José Antonio. «Tesauros: Estándares y recomendaciones» [en línia]. *El Profesional de la Información*. Núm. 21 (maig-juny del 2012). <<http://www.elprofesionaldelainformacion.com/contenidos/2012/mayo/02.pdf>> [Consulta: 23 de desembre de 2016].
- › SCHREIBER, Guus; RAIMOND, Yves. *RDF 1.1 Primer. W3C Working Group Note 25 February 2014* [en línia]. <<https://www.w3.org/TR/2014/NOTE-rdf11-primer-20140225/>> [Consulta: 3 de desembre de 2016].
- › SHEARER, Kathleen. *Confederation of Open Access Repositories. Promoting Open Knowledge and Open Science Report of the Current State of Repositories. May 21, 2015* [en línia]. COAR, 2015. <<https://www.coar-repositories.org/files/COAR-State-of-Repositories-May-2015-final.pdf>> [Consulta: 9 de desembre de 2016].
- › SULÉ, Andreu; CENTELLES, Miquel; FRANGANILLO, Jorge; GASCÓN, Jesús. «Aplicación del modelo de datos RDF en las colecciones digitales de bibliotecas, archivos y museos en España» [en línia]. *Revista Española de Documentación Científica*. Vol. 39. Núm. 1 (2016). <<http://redc.revistas.csic.es/index.php/redc/article/view/924/1340>> [Consulta: 12 de gener de 2017].
- › 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*. Vol. 24 (març-abril del 2015). <<http://www.elprofesional-delainformacion.com/contenidos/2015/mar/04.html>> [Consulta: 15 de desembre de 2016].
- › TSALAPATIL, Eleni; SIMOUL, Nikolaus; DROSOPOULOS, Nasos; STEIN, Regine. «Envolving LIDO based aggregations into Linked Data» [en línia]. Dins: *CIDOC Annual Conference 2012. Enriching Cultural Heritage*, Hèlsinki, 10-14 de juny de 2012. <http://network.icom.museum/fileadmin/user_upload/minisites/cidoc/ConferencePapers/2012/simou.pdf> [Consulta: 16 de desembre de 2016].

- › TZOMPANAKI, Katerina; DOERR, Martin. *Fundamental Categories and Relationships for intuitive querying CIDOC-CRM based repositories*. *Technical Report ICS-Forth/TR-429*. April 2012 [en línia]. <http://www.cidoc-crm.org/sites/default/files/TechnicalReport429_April2012.pdf> [Consulta: 11 de desembre de 2016].
- › TZOMPANAKI, Katerina; DOERR, Martin. [2013]. *Reasoning based on property propagation on CIDOC-CRM and CRMdig based repositories* [en línia]. <<http://ceur-ws.org/Vol-1117/paper4.pdf>> [Consulta: 11 de desembre de 2016].
- › VALIAKIS, Konstantinos; KARAGIANNIS, Georgios Th.; MIKAS, Pericles A. «Semantic Web in Cultural Heritage After 2020» [en línia]. Dins: *What Will The Semantic Web Look Like 10 Years From Now? In conjunction with the 11th International Semantic Web Conference*. Boston, 11-15 de novembre de 2012. http://stko.geog.ucsb.edu/sw2022/sw2022_paper10.pdf [Consulta: 30 de novembre de 2016].
- › VAN HOOLAND, Seth; VERBORGH, Ruben. *Linked Data for Libraries, Archives and Museums. How to Clean, Link and Publish Your Metadata*. Chicago: Neal-Schuman, 2014. P. 43-44.
- › VILA SUERO, Daniel; VILLAZÓN TERRAZAS, Boris; GÓMEZ PÉREZ, Asunción. *datos.bne.es: a Library Linked Data Dataset* [En línia]. 2012. P. 1-6. <<http://www.semantic-web-journal.net/sites/default/files/28620-Vila-Suero.pdf>> [Consulta: 5 de març de 2017].
- › VILLAZÓN TERRAZAS, Boris; VILCHES BLÁZQUEZ, Luis M.; CORCHO, Óscar; GÓMEZ PÉREZ, Asunción. *Methodological Guidelines for Publishing Government Linked Data* [en línia]. 23 p. [2011]. <https://www.lri.fr/~hamdi/datalift/tuto_inspire_2012/Suggestedreadings/egovld.pdf> [Consulta: 11 de febrer de 2017].
- › ZAHAREE, Marcie. «Building controlled vocabularies for metadata harmonization» [en línia]. *Bulletin of American Society for Information Science and Technology*. Vol. 39. Núm. 2 (2012/2013) <http://www.asis.org/Bulletin/Dec-12/DecJan13_Zaharee.pdf> [Consulta: 3 de març de 2017].
- › QUÍLEZ MATA, Julio: “Aproximació a les ontologies. Definició i construcció. Aplicacions en el camp de l'arxivística” [En línia]. Lligall: *Revista catalana d'arxivística* Núm. 32 (2011). P. 105-156. <<https://www.arxiv.org/index.php/documents/publicacions/revista-lligall-1/lligall-32-1/998-05-aproximacio-a-les-ontologies-definicio-i-construccio-aplicacio-al-camp-de-l-arxivistica-1/file>> [Consulta: 20/01/2017].
- › BRIATTE, Katell; HADOC. *Modèle harmonisé pour la production des données culturelles*. Ministère de la culture et de la Communication. 2013 [En línia] <<http://www.culturecommunication.gouv.fr/Divers/Harmonisation-des-donnees-culturelles/Modele-de-donnees/Enjeux-et-objectifs>> [Consulta: 21/02/2017]
- › SIBILE – DE GRIMOÛARD, Claire; MOTTE, Alice: « Archives and Linked Data: are our tools ready to “complete the picture”? Two case studies » [en línia]. Dins: *Arxius i Indústries Culturals*, 11-15 octubre de 2014. Girona. P. 1-11 <<http://www.girona.cat/web/ica2014/ponents/textos/id9.pdf>> [Consulta: 08/01/2017]

RESUM

La necessitat d'integrar i fer accessible el contingut de les col·leccions d'arxius, biblioteques i museus ha convertit aquestes institucions i el patrimoni cultural en conjunt en el camp d'experimentació de les tecnologies del web semàntic, atesa la riquesa dels seus fons, la multidisciplinarietat en el seu tractament i l'experiència dels seus professionals, especialment de l'àmbit de les biblioteques, en el terreny de la creació, la gestió i el manteniment de metadades. Aquest article introdueix les principals línies d'acció d'una de les estratègies de desplegament del web semàntic més reeixides dels darrers temps: linked open data.

Paraules clau: *linked open data, patrimoni cultural, arxius, biblioteques i museus, web semàntic, ontologies, CIDOC CRM, dades obertes, reutilització de la informació, transparència de les administracions, Europeana.*

RESUMEN

La necesidad de integrar y hacer accesible el contenido de las colecciones de archivos, bibliotecas y museos ha convertido a estas instituciones y a su patrimonio cultural en el campo de experimentación de las tecnologías de la web semántica, dada la riqueza de sus fondos, la multidisciplinariedad en su tratamiento y la experiencia de sus profesionales, especialmente aquellos del ámbito de las bibliotecas, en el terreno de la creación, la gestión y el mantenimiento de metadatos. Este artículo introduce las principales líneas de acción de una de las estrategias de desarrollo de la web semántica más exitosas de los últimos tiempos: linked open data.

Palabras clave: *Linked open data, patrimonio cultural, archivos, bibliotecas y museos, web semántica, ontología, CIDOC, CRM, datos abiertos, reutilización de la información, transparencia de las administraciones, Europeana.*

ABSTRACT

The need to integrate and make accessible the contents of archives, libraries and museums has made these institutions, and cultural heritage in general, the experimental field of Semantic Web technologies, due to their wealth of holdings, multidisciplinary of treatment and the experience of professionals, particularly in libraries, in the creation, management and maintenance of metadata. This article introduces the main lines of action of one of the most successful strategies of Semantic Web development of recent times: linked open data.

Keywords: Linked Open Data, Cultural heritage, archives, libraries and museums, Semantic Web, ontologies, CIDOC, CRM, open data, re-use information, transparency of government, Europeana.

RÉSUMÉ

La nécessité d'intégrer et de rendre le contenu des collections des archives, des bibliothèques et des musées accessible a fait de ces institutions, ainsi que du patrimoine culturel dans son ensemble, un champ d'expérimentation des technologies du web sémantique, en raison de la richesse de leurs fonds, de la multidisciplinarité dans leur traitement et de l'expérience de leurs professionnels, notamment ceux des bibliothèques, dans le domaine de la création, de la gestion et de la maintenance des métadonnées. Cet article introduit les principales lignes d'action d'une des stratégies de déploiement du web sémantique les plus réussies à l'heure actuelle : le linked open data.

Mots-clés: Linked Open Data, patrimoine culturel, archives, bibliothèques et musées, Web sémantique, ontologies, CIDOC, CRM, les données ouvertes, la réutilisation des informations, la transparence du gouvernement, Europeana.