



VII. Arte Garaikideko Dokumentazio Zentroen Topaketak

Artiumeko Liburutegi eta Dokumentazio-Zentroa

2014ko urriaren 29an eta 30ean. Artium Museoa
Informazio-zerbitzuen lidergoa XXI. mendean

LOD-LAM erresuman ... Linked Open Data eta Ondare Digitala

Hitzaldiak: Eva M. Méndez Rodríguez eta Gema Bueno de
la Fuente

Bibliotekonomia eta Dokumentazioa Carlos III Unibertsitatea,
Madril

LOD-LAM erresuman... Linked Open Data eta Ondare Digitala

1. Sarrera

«XXI. mendeko informazio-zerbitzuen lidergoaz» mintzatzeko, behar-beharrezkoa da zerbitzu horiek modu nabarmenean egotea sarean sartuta. 25 urte badira Web sarea dugula –eta zenbaitentzat, badira 15 urte Web edo sare semantikoa dugula–, eta World Wide Web ez da jadanik «egon behar den tokia», baizik eta «modu jakin batean egon behar den tokia», alegia, «izan behar den tokia». Web izatea da informazio-zerbitzuentzat izan daitekeen gauzarik onena. Halako zerbitzutzat ulertzen ditugu era guztietako liburutegiak, artxiboak, museoak eta galeriak, eta esan behar da azkenaldi honetan ontzat eman dutela ingelesezko siglez osaturiko zorioneko izendapenari erantzutea: LAM (Libraries, Archives and Museums) edo GLAM, galeriak ere sartzen baditugu.

Sekula ez dugu izan orain bezain gertu kultura-ondare globala. Azken hamar urteotan eginahal handiak egin dira eta dirua erruz inbertitu da, bai erakunde publikoena, bai erakunde pribatuena, bildumen eta edukien digitalizazioan, hala nola kultura-ondare berriaren bilduma eta objektu digital horiek deskribatzen dituzten metadatuak erregistroan. Alabaina, lan handia egin behar da oraindik bilduma horien interoperabilitatera iristeko, bilaketa soilak egiteko eta hainbat bildumatan eta informazio-zerbitzu digitaletan barna informazioa modu garden, integratu eta eraginkorrean berreskuratzeko. Memoriaren erakundeek, GLAM hauek, esperientzia soziala nahi dituzte beren bilduma digitalizatuak bihurtu, edukien erakusketa baino areago. Funtsezkoa da bildumak, erakustekoak eta ikustekoak izatetik, berrerabiliak izatera pasatu ahal izatea, hau da, museotik, liburutegitik edo artxibotik kanpora ateratzea.

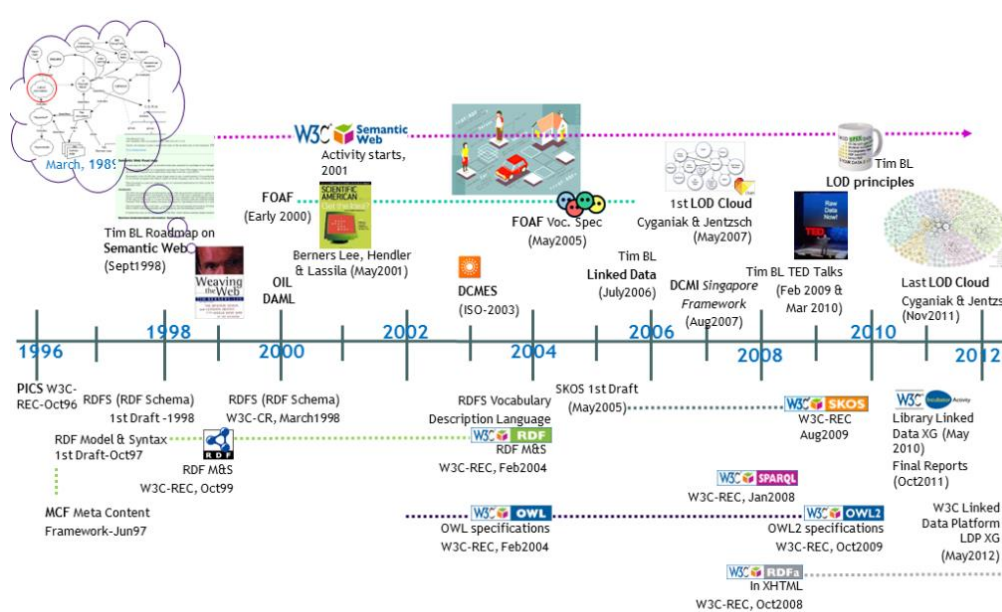
LOD edo Linked Open Dataz hitz egitea ohikoa da dagoeneko Web informazio-testuinguru batean, baina paradigma berri batez, espazio berri batez, erresuma berri batez nahi genuke hemen hitz egin. Aspalditik ari dira GLAMak metadatuak sortzen, interoperabilitatea hobetzeko agindu isil eta prestuarekin; interoperabilitatea, askotan zer esan nahi duen oso ondo ez dakigun hitz magiko hori, sareko informazio-zerbitzu desberdinen artean. Alabaina, ez da beti guk nahi genukeena bezalakoa metadatu horien kalitatea, erakusketa- edo erabilgarritasun-maila. Teknologia, estandar eta protokolo desberdinen inplementazioak sakabanatu egin ditu gure museoetako, liburutegietako edo artxiboetako metadatuak, datu-base, kalkulu-orri eta baita testu-dokumentu egituratugabeetan ere. Bilduma horietako askoko metadatuak xehetasun eta kalitate maila desberdinak dituzte, eta datu-siloak eta uharte semantikoak (semantikoak, baina uharteak azken batean) osatzen dituzte maiz. Linked Data (LD) izenekoak agindu berri bat suposatzen du, hau da, bilduma desberdinetako objektuen artean estekak edo loturak sortzea, nahiz eta inplementazio errealak oraindik ere halako konplexutasun bat izan, eta nahitaez ulertu beharra dago nola funtzionatzen duten web semantikoen estandarrek (RDF, SPARQL, etab.), horiei esker dimentsio «estekatu» berri horretaz hornitu ahal izango baititugu gure informazio digitaleko objektuak.

Zailtasun horrengatik guztiagatik ere, eta LOD erresuma datu-nahasketa handi korapilotsu bat bihur daitekeen arren, informazio-zerbitzuentzako eta kultura-ondare digitalizatuentzako erronka berria delakoan gustatzen zaigu Linked Open Dataz hitz egitea, top-down aldaketa bat delakoan, dagoeneko asko baitira joera berri honi heldu dioten memoriaren erakundeak. Hala bada, liburutegi, artxibo eta museoetako LODen erabilera eta aplikazioen kasu erreal batzuk analizatuko ditugu argitalpen honetan, baita memoriaren erakundeen dataset izenekoak –silo edo uharte semantiko partikularretan jasoak gaur egun– LOD galaxiara igarotzen lagunduko duten printzipio eta oinarri batzuk ere, eskuragarri egon daitezten eta berriz erabiltzeko moduan Sektore Publikoaren Informazioaren (SPI) funtsezko parte gisa.

2. Web semantikoaren bilakaera eta heldutasuna

Lehen esaten genuen bezala, sarea duela 25 urte existitzen da; duela 20 urte Dublin Core-a eta metadatuak. Webari semantika ipintzea ere ez da gauza berria, 15 urtetik gora pasatu baitira RDF estandarra eztabaidatzen (Resource Description Framework) hasi zenetik. Hamabost urte askotxo dira web semantikoaz etorkizuneko weba delakoan hizketan jarraitzeko. Garai batean informazio-zerbitzu digitalentzako agindu bat zena, «web bibliotekarioago» batera eramango gintuena gainera (Méndez, 2004), berehalako prozesu bat da gaur egun: dokumentuen web-etik, datuen web-era igarotzea.

Datu estekatuen web-aren garapena lortzeko, datuak formatu bateratu batean (RDF, Resource Description Framework) eskuragarri egotea bideratu behar dute teknologiek, formatu horretan zuzenean sortuak edo lehendik zeuden datu-baseetatik abiatuta bihurtuta (erlazionalak, XML, HTML, etab.).

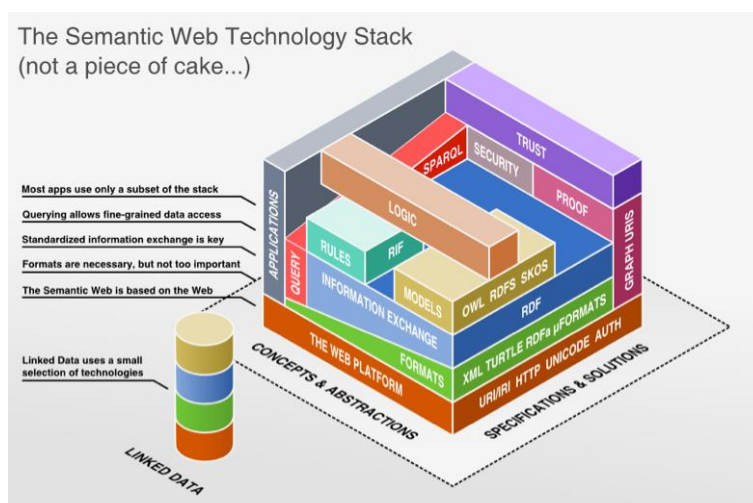


1. irudia. Web semantikoaren 15 urte: estandarren bilakaera (Méndez, Greenberg, 2012)

3. Linked Open Dataren oinarriak: estandarrak eta beste bidaia-lagun batzuk

3.1. Web Semantikoarentzako estandarrak

Aurreko atalean argi utzi dugu Webaren bilakaera, zentzu batean, bere estandarren bilakaera dela, eta, batez ere, Linked Open Data eta web semantikoa oinarritzen diren zutarriak oraindik ere estandar formalak izateko bokazioa duten herri eskupekoko estandarren edo espezifikazioen multzo bat direla. Kasurik gehienetan W3Cren Gomendioak direlakoan daude pentsatuak. RDF, GRDDL, POWDER, RDFa, R2RML, RIF, SPARQL: horra hor sigla batzuk web hau osatzen duten datuetarako sarbidea eta argitaratzeko aukera ematen duen arau tekniko eta teknologien multzoa adierazten dutenak.



2. irudia. Web semantikoaren teknologien multzoaren irudikapen grafikoa

(Iturria: http://bnode.org/media/2009/07/08/semantic_web_technology_stack.png)

Teknologia hauek hainbat geruzako pakete baten moduan irudikatzen dira gehienetan (2. irud.), eta ondorengo hauei dagozkien kontzeptu eta abstrakzioen inguruan antolatzen dira hartan estandarrak edo espezifikazioak: web plataforma edo arkitektura, formatuak, informazio-trukea, datuen ereduak, kontsulta eta lokalizatorako funtsezkoak diren beste alderdi batzuk, datuen berreskuratzea eta haietarako sarbidea inferentzia-arauen definizioaren bitartez, logika, segurtasuna, babesa eta konfiantza.

Hona hemen espezifikazio garrantzitsuenetako batzuk:

- *Web plataforma/arkitektura mailan*: URLak (Uniform Resource Identifiers) edo IRIak (Internationalized Resource Identifiers), IETF (Internet Engineering Task Force) delakoaren baliabideen identifikatzailearentzako estandarrak. Baliabideen identifikazio jarraitua da, datuen sorkuntza-, bihurtze-, argitalpen-, lokalizazio-, eta kontsumo-prozesu osoan, Datuen Webaren eta LODen

paradigman oraindik ere garrantzi gehiago hartzen duen alderdi bat. URI horiek, gainera, deferentziagarriak izan behar dute, eta, horretarako, webaren beste oinarritzko estandar bat erabili behar dute, HTTP protokoloa (HyperText Transfer Protocol). HTTP URLak datuen izentzat hartzen dira, ez web-helbidetzat.

- *Formatuei* dagokienez: markatzeko hizkerak, esate baterako XML (eXtensible Markup Language) edota baliabideak atribuetan deskribatzen dituen esparrua, RDFa (Resource Description Framework—in—attributes) datuak webeko dokumentuetan egituratzen eta txertatzen dituzten formatuak dira.
- Informazioaren *trukeri* dagokionez: RDF (Resource Description Framework) baliabideen deskribapen-esparrua dela medio, datuen irudikapena, pertsonentzat ez ezik, makinentzat ere modu esanguratsu eta ulergarrian estandariza daiteke, eta era horretan posible da plataformen eta sistemen arteko datuen multzoa trukitzea. RDF datu-multzo hauek hirukoitz edo tripletan biltegitratzen dira, triple store izenekoetan, haien kontsulta edo/eta deskarga egin ahal izateko.
- *Kontsultarako*: beharrezkoa da kontsulta komunerako hizkera bat, eta RDF tripleen biltegien kontsultak erregulatzen dituen protokolo bat, eta, horretarako, SPARQL (Simple Protocol and RDF Query Language) garatu da.
- *Eta datuen modelatua* egiteko: ontologiaren definiziorako espezifikazioek, nola den OWL (Ontology Web Language), eta jakintzaren beste antolamendu-sistemek edo hiztegiak, SKOS izenekoak esate baterako (Simple Knowledge Organization System), datuen modelatua ahalbidetzen dute, datuen eta propietateen definizioaren bitartez, hala nola horien arteko harremanen definizioaren bitartez.

3.2. Metadatuaren paradigma

Liburutegiaren eta ondare digitalaren komunitateak eginahal handi bat egin du azken 20 urteotan metadatu eta hiztegi semantikoaren eskema normalizatzen, hala nola metadatu-bilduma ikaragarri handiak sortzen. Ekar dezagun gogora, esate baterako, Europeana, zeinak gaur egun (2014), ia 36 milioi metadatu-erregistro jasotzen baititu, Europako memoriaren erakunde desberdinen beste horrenbeste informazio-objektu digitalizatu deskribatzen, eta nolakotasun desberdineko hiztegiaren kodifikazio-eskema baliatzen dituztenak (tesauruak, ontologiak, etab.).

Seth van Holland-ek eta Ruben Verborg-ek duela gutxi kaleratu duten liburu batean (2014) azaltzen dutenez, egituratu gabeko informazio batetik oso egituratuak dauden metadatuaren irudikapenerako bilakaerak schemas garatzea behar du, metadatu horiek interoperagarriak izan daitezten. Egituratu gabeko baliabideak alor ondo egituratuak bihurtzeko, alor horien (atributuen) semantikak eskema batean ondo dokumentatua behar du egon, eta horretan saiatu behar dugu. Metadatuaren alorrak, egituratuz eta atomizatuz, makinentzako interoperagarriagoak egiten ditugu, baina berebat egiten ditugu eskemaren gero eta mendekoagoak. Gainera, inguru digitalean garrantzia izan nahi duen komunitate, domeinu edo proiektu bakoitzak bere metadatuaren eskema, bere hiztegia edo hiztegiak garatzen ditu, edo, hala ez bada, aurretik zegoen eskema baten egokitzapena, zeina baita aplikazio-profil gisa ezagutzen duguna. Era horretan, gero eta zailagoa gertatzen zaigu informazio digitaleko beste proiektu edo sistema baten metadatuak interpretatzea, ez badugu haren schema ezagutzen. Hori dela-eta, metadatuak askotan hortzetako eskuilak bezalakoak dira: gauza ona direla pentsatzen dute denek, baina bakoitzak berea erabili nahi du.

Gero eta gehiago dira, GLAMen baitan eta kanpoan, metadatu-eskemak (schemas), Dublin Core esate baterako, sortzen eta erabiltzen dituzten komunitateak, edota metadatu partikularren estandarrak eta aplikazio-profilak. Talde edo komunitate hauek beren estandar deskribatzaileak sortzen dituzte, zeinek elementuen eta propietateen (atributuen) multzoak jasotzen baitituzte, beren metadatu-hiztegiak osatzeko. Era berean, W3Caren web semantikoaren komunitateak, duela urtebete baino gutxiago hasita, datuen komunitatean murgilduta, esan dugun bezala, ontologia eta hiztegi formalizatuak garatzen jarraitzen du. OWL (Web Ontology Language) erabiltzen dute hiztegi horiek eta ontologiak eraikitzeke, eta SKOS jakintzaren antolamendu-sistemak (KOS) sortzeko, datuen beharrezko aberastearen prozesu konplexu baten parte gisa informazio-inguru berrian. Web-garatzzaileek berek ere webguneekin lotutako metadatuak berpiztu dituzte ostera ere, HTML dokumentuen edukia deskribatzen edo idazten dituzten datu egituratuak erantsiz. html dokumentu baten kode-iturrian txertaturiko deskribapen hauentzat, kodetze-estandarrak erabiltzen dituzte, hau da, mikroformatuak, mikrodatuak eta RDFa, eta berebat erabiltzen dituzte beren hiztegia edo beren metadatu partekatuak, Schema.org adibidez.

Metadatuaren eskemen eta beste hiztegien sorrerak, nolana ere, informazio digitalaren kudeaketaren parte dira, eta hiztegi hauek guztiek ohiko paradigma bat osatzen dute jadanik.

3.3.LOD paradigma: Linked Open Dataren printzipioak eta oinarriak

Azken 20 urteotan, teknologiei, edo, hobeto esanda, Interneti esker, oso maiz aditzen dugu, maiztxo agian gure iritziz, «paradigma berri baten aurrean» gaudela, edota halako eta bestelako jokamolde teknologikoak edo teknologiak paradigma-aldaketa bat suposatzen duela. Alabaina, paradigma baten berezitasun nagusia, ikuspegi epistemologiko batetik, haren oinarri-balioa da, oinarriak emateko nolakotasuna izatea alegia, horrek sortzen baitu esparru teoriko edo teoria-multzo berria. Hala, datu ireki estekatuaren eskakizunek edo Linked Open Data (LOD) izenekoek, Tim Berners-Leek bere bost izarreko eskema inkrementalean definitzen dituen moduan (Berners-Lee, 2006), era horretan formulatu gabe (paradigma berri bat), beharrezko baldintzak zertzen ditu datuek eredu berri bati erantzun diezaioten. Hala, LODi buruz hitz egin ahal izateko, datuek sarean argitaratuak behar dute egon, alegia, URI bat behar dute izan, makinak irakurtzeko moduan egituratuak egon behar dute, formatu ez-jabeetan, W3Caren estandarren arabera, eta beste datu-multzoekin estekaturik. Baldintza horiek, gainera, modu teknikoagoan esplizitatzen dira Linked Dataren lau printzipioen gisa (Heath eta Bizer, 2011):

1. URIak erabili gauzen izentzat.
2. URI HTTPak erabili, jendeak izen horiek bilatzeko eta lokalizatzeko moduan.
3. Norbait bilaketan dabilenean URI bat erabiliz, informazio erabilgarria eskaini, estandarrak baliatuz (RDF, SPARQL).
4. Beste URIlekiko estekak baliatu, gauza gehiago aurkitu eta lokalizatu ahal izateko.

4. GLAM erakundeak eta informazioaren berrerabilpena sektore publikoan

Hasieran, 2010. urtean, LOD-LAM aipatu ohi zen (<http://lodlam.net>) Linked Open Datak oroimenaren erakundeetan (liburutegi, artxibo eta museoetan) izan zitzakeen aplikazioaz hitz egiteko, Jon Voss-i jarraituz. Harrezkero, zaleek, teknikariek, profesionalek eta, funtsean, «freakiek» osaturiko sare bat dela pentsatzen da, zeinak interesatuak baitaude galeria, liburutegi, artxibo eta museoekin zerikusia duten LODekin lan egiten. Hori dela-eta, GLAMez ere hitz egiten da Galeriez hitz egitean. Guretzat, inguru ezin hobea izan dira beti GLAMak Linked Open Datarentzat. Linked kontzeptuak lotzeaz dihardu, W3Caren estandarrak erabiliz horretarako, eta Open, berriz, oroimenaren erakundeen edukientzako lizentzia irekien erabileraz ari da, esate baterako CC0 (Creative Commons Public Domain Mark, etab.). Eta datuak dira, nagusiki, Tim Berners-Leeren «raw data» izenekoak, edota metadatuak bestela, datu deskribatzaileak, datu bibliografikoak, hiztegiak eta beste datu batzuk, zeinak, LOD-[G]LAM erresuman, aise, baina halabeharrez, izan bailitezke irekiak, estekatuak eta partekatuak. Guretzat hori hala bada ere, eta errealitate bat den arren LODLAM komunitatearentzat, 2011n antolatu baitzuen bere lehen workshopa, itxoin egin beharko genuen GLAMak sektore publikoaren informazioaren parte gisa onetsi ziren arte, ireki daitekeena, eta, areago, estekaz lotu daitekeena, paradigma berriaren parametroen arabera.

Hona nola ulertzen den informazioaren berrerabilpena sektore publikoan: «Pertsona fisikoek nahiz juridikoek sektore publikoko erakundeek sorturiko informazioa erabiltzea, merkataritza-xedeekin edo gabe» (Aporta proiektua, 2009). Hori hortxe dago, eta gardentasunarekin eta «publikoaren» nozioarekin dago lotua. Alabaina, garapen teknologikoak bultzatu edo areagotu du berrerabilpen hori, ez bakarrik IKT izenekoek eskaintzen dituzten sarbide- eta truke-aukerei dagokienez, baizik eta, bereziki, datuen analisisan, ustiaketan eta tratabidean izan duten bilakaeragatik, horri esker sor baitaitezke datuen erabileran, eransketan edo konbinazioan oinarrituriko zerbitzuak eta aplikazioak.

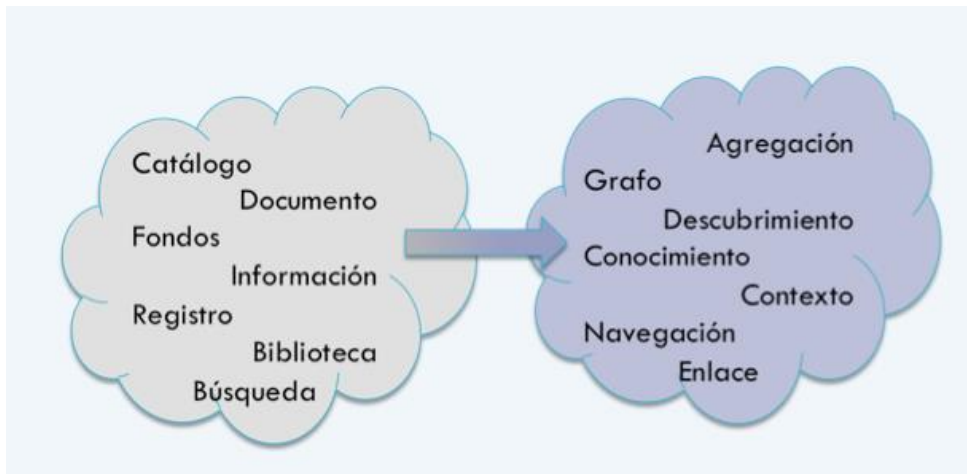
Berrerabilpen hori administrazio publikoek sustatzen dute, Europar Legebiltzarraren beraren eta Batzordearen aholkuz, bere Zuzentarau 2003/98/CEarekin kultura-aldaketa bat nahi baitute haiek katalizatu, informazio kopuru handi eta zabal bat berrerabiltzeko xedez, fidagarria eta kalitatezkoa, erakunde publikoek ekoizten eta kudeatzen dutena. Espainian, bultzada hori 37/2007ko Legea, azaroaren 16koa, sektore publikoko informazioa berrerabiltzeari buruz izenekoan gauzatu da, zeina 1495/2011 Errege Dekretuan garatu baita, eta, beste emaitzen artean, datu irekien atarrien sorrera suposatu baitu, nola den datos.gov.es estatu mailan, eta baita beste batzuk ere, zenbait autonomia- elkargoren gobernuen, udaletxeen eta beste erakunde batzuen ekimenez.

Europar Batzordeak zuzentarau hau berrikusteko ekimena bideratzea proposatu zuelarik, lehendabizi 2011n (Europar Batzordea, 2011), eta 2013an gero, osatu egin da aurreneko bertsioan oroimenaren erakundeak baztertzeko suposatzen zuen hutsunea, eta liburutegietara (unibertsitateko liburutegiak barne), artxiboetara eta museoetara zabaldu du bere aplikazio-eremua. Aitortzen da era horretan artxiboko materialak, liburu-bildumak eta artelanak ere erakunde publikoen informazioaren eta edukien ondarearen parte direla. Konpromiso publikoa duten liburutegiaren unibertsoaren konplexutasunak eta aberastasunak berrerabilgarriak izan daitezkeen bildumen eta datuen aniztasun eta tipología zabala dakarte. ¹

Era berean, Europar Batzordeak, 2014ko uztailean, hiritarrak datu-ekonomia oparo batera bultzatzen dituen Komunikazio bat argitaratu zuen. Bertan honela definitzen dira datuak, ISO/IEC 2382-1-en arabera: «Informazioaren irudikapen berrinterpretagarria, modu formalizatuan, eta komunikaziorako, interpretaziorako edo prozesatzeko modu egokian egina» (Europar Batzordea, 2014). Hona hemen adibide batzuk: informazio geoespaziala, estatistikak, datu meteorologikoak, ikerketa-datuak, etab.; baita gure datu bibliografikoak ere, eta bestelako nolakotasuna duten datu guztiak, zeinak, gero eta gehiago, prozesatzen, komunikatzen, irudikatzen edo berrerabiltzen baitira, GLAM informazio-sistema eta -zerbitzuetan. Horren adibide batzuk ikus daitezke W3Cko Library Linked Data Incubator Group-ek aztertu dituen erabilera-kasuetan (W3C Incubator Group, 2012).

Badira hamarraldi batzuk liburutegiaren paradigma-aldaketa aipatzen dela, paperezko tradizioetik liburutegi digitalera, atomotik bitera; baina oraintxe ari da, inoiz baino gehiago, trantsizio hori errealitate bihurtzen. Stefan Gradmannek bere *From Containers, to Content, to Context* txostenean dioen bezala (Gradmann, 2012), liburutegiek, edukitzaileak izatetik, informazioaren testuinguru-sortzaile izatera behar dute igaro. Aukera hori ezin da galdu, ez badugu garrantzi gabeak izatera heldu nahi. Gradmann, trantsizio horren puntu nagusiak aipatzen dituenean, terminologia aldatzeko premiatik hasten da (3. irud).

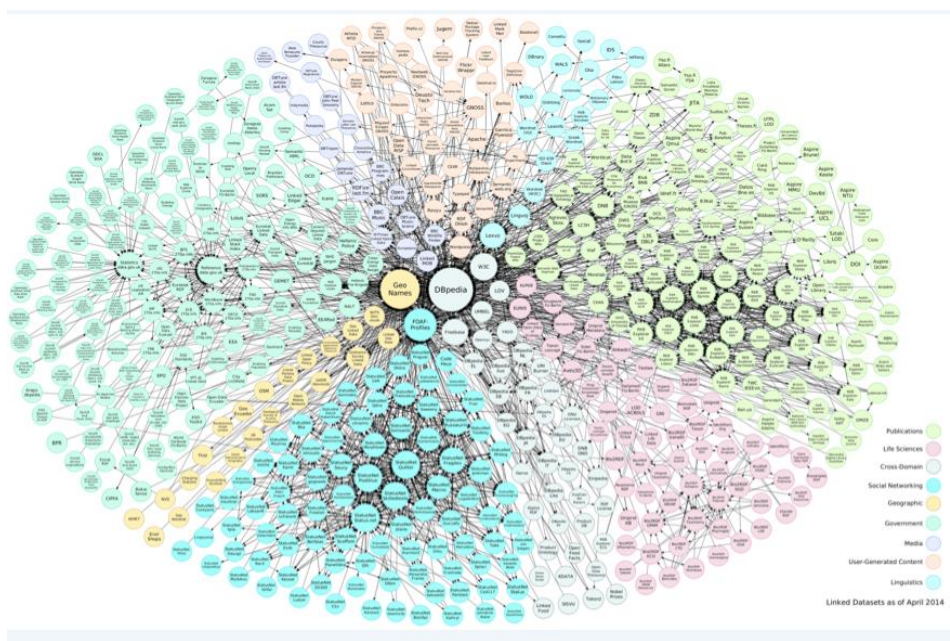
¹ Azpimarratu behar da, nolana ere, ezin direla liburutegiaren funts guztiak ireki eta berrerabili, legedia ez baitaie aplikatzen jabetza-eskubide intelektualak dauzkaten dokumentu eta datuei. Hori dela-eta, berrikusi behar dela esan izan da 2013ko ekainaren 26an argitaraturiko zuzentaraua (Europar Legebiltzarra, Batzordea, 2013).



3.irudia. Terminologia-aldaketa liburutegietan (itzulpena Stefan Gradmann, 2012, 41 or.)

5. Uharte semantikoetatik Linked Data Cloud-era

Linked Data Cloud-en (4 irud.) 570 dataset edo datu-multzo daude gaur egun jasoak, eta horietatik 374 hornitzaileek berek deskribatu zituzten datahub.io dataseten katalogoan, eta 196 datu-multzo identifikatu ziren jarraipen automatiko bat eginez LOD Webgunean, 2014ko apirilean.



4. irudia. Linked Data Cloud-en diagrama (Iturria: <http://data.dws.informatik.uni-mannheim.de/lodcloud/2014>)

Hona hemen bururatzen zaigun galdera nagusia, batez ere, hodeia «hain polita» delarik, baina hain gutxi ikusten da... Zer dago jende guztiak datuen, datu-baseen eta gauza guztien paradigmaren aldaketa iradokitzeke erabiltzen duen LOD hodei horretan? LOD printzipioen arabera sarean argitaratzen diren eta beste datu-multzoekin estekatzen diren datu-multzoak erakusten ditu LOD hodeiaren diagramak. Zirkuluak datu-multzoak dira, eta koloreak sailkatzeko ezarri diren domeinuei dagozkie: argitalpenak, bizitzaren zientziak, gizarte-sareak, datu geografikoak, gobernuak, bitartekoak, erabiltzaileak sorturiko edukia, hizkuntzalaritza eta domeinuarekin zerikusia ez duten beste batzuk. Geziek bi RDF datu-multzoen arteko estekak adierazten dituzte, eta zirkuluen tamaina datu-multzo bakoitzak beste datu-multzo batzuei dagokienez duen esteka-graduaren mendekoa da. Hodeia dela medio, erraz identifika daitezke, bai hodei osoaren erdian (DBpedia, Geonameak, etab.), bai domeinu bakoitzean, beste iturrietatik datozen esteka-kopururik handiena kontzentratzen duten datu-multzoak.

Gero eta gehiago dira LOD sare honetan dauden oroimenaren erakundeak, eta datu-multzo desberdinak eskaintzen dituzte, besteak beste, hiztegi-kopuru eder bat, Agrovoc bezalako tesauruak, materia-goiburuak (Library of Congress-ko LCSH delakoa, baina baita Espainiako liburutegi publikokoak ere), DDC bezalako sailkapenak, autoritate-zerrendak, bibliografia-baliabideak, bilduma-katalogoak, baliabide-zerrendak, bilaketarako laguntzak, liburutegietako, artxiboetako eta museoetako direktorioak, besteren artean.

6. LODen erabilera GLAMen munduan: kasu batzuk

Kutsatzea izaten da gauza bat zabaltzeko modurik hobereana. Hala, LOD teknologiak oroimenaren erakundeetan ari dira zabaltzen. Hasieran, liburutegiek bereganatu zituzten (Byrne eta Goddard, 2010, eta Baker et al., 2012), baina hasi dira berebat aurreko atalean aipatu dugun Linked Data Clouden parte izaten, hau da, artxiboetako, museoetako eta galerietako datu-multzok osatzen, OpenGLAM joera osorako espazio interesgarri bat taxutuz. Datuak, metadatuak, lotu edo estekatu nahi diren «erresuma» berri horretan, hortxe behar dugu datuen ulerpen hobe batera heldu. Datu-eredu soil bat hartuz, abstrakzioa egin daiteke XML eskemetatik eta datu-base tradizionaletatik, bibliografia-katalogoetatik eta abarretatik abiatuta, zeinak iraganean erabili baitziren datuak interpretatzeko eta erabiltzeko, eta, orain, berrerabiltzeko asmoa har genezake.

Ondoren datuen, metadatuaren eta informazioaren erakustaldi-zerbitzuen kasu enblematiko batzuk aurkeztuko ditugu, beti ere LOD-LAM paradigmaren baitan, zeinei, orain gutxi (2014ko azaroa), Espainiako Liburutegi Nazionalako katalogo berria erantsi baitzaio (datos.bne.es). Duela hamabost urte aukera gisa aipatzen genuen hura (Méndez, 1999), errealitate bihurtu da, eta RDF da katalogazioaren eta metadatuak erakusteko moduaren beste kodifikazio posibleen alternatiba; ez bakarrik informazioaren berreskuratzea hobetzeko, baita ondare kulturalaren informazio-aktiboak deskribatzen dituzten datuen berrerabilpenari bidea zabaltzeko ere.

6.1. Europeana: ondare kulturalaren LOD pilotua

Alabantza askorekin, eta beste horrenbeste kritikarekin, Europeana bihurtu da, beste guztien gainetik, Europako ondare kulturalari sarbidea ematen dion baliabidea, zerbitzua edo plataforma. Hasieran Europar Liburutegi Digitala zena, denborarekin, «liburutegitik gutxiago eta digitaletik gehiago» izatera heldu da (Méndez, 2010) Europako ondare digitalizatu aberatsera sarbidea bultzatzeko, eta, baita ere, gero eta gehiago, sormena, berrikuntza eta berrerabilpena sustatzeko xedez.

Linked Open Data da Europeanaren edukiaren berrerabilpena indartzeko moduetako bat. data.europeana.eu, datu ireki eta estekatuen pilotua (Europeanaren kasuan, metadatuak), 2010. urtean jarri zen abian, EuropeanaConnect eta Europeana v2 proiektuen baitan, beren metadatuak LOD gisa sarean jartzeko eginahal adostu batean. Horren bitartez, Europeanak metaturiko eta zenbait datu-hornitzaileengandik heldutako metadatuetarako sarbidea bideratuko litzateke, web teknologien bitartez (RDF, SPARQL, etab.), eta baita metadatu horiek berriz erabili eta hornitzaileari aberasturik itzuliko ere.

Era honetan egin daitezkeen estekak direla medio, Europeanaren baliabideetatik (LOD gisa erakutsitako bere metadatuaren bitartez) lehenago erakutsi dugun Linked Dataren Webguneko beste baliabideekin (4. irud.), orobat berreskuratu ahal izango dira semantikoki lotuak eta berebat irekiak eta estekatuak dauden beste baliabide batzuk; hala, esate baterako, artista beraren bi artelan estekatzeko aukera sortuko litzateke, **loturaren bat badute, nahiz eta bilduma (silo) digital berekoak izan ez.**

6.2. British Museum: Bilduma Semantikoa


British Museum izan da bere datuak LOD hodeian ipini dituen munduko lehen museoetako bat. Zehatzago, sarean dauka bere katalogoa, museoko 3,5 milioi objektu deskribatzen dituzten 2 milioi erregistro baino gehiagorekin, eta CIDOC CRM ontologiaren bitartez irudikatu da (Conceptual Reference Model). ResearchSpace proiektuaren emaitza gisa, LOD printzipioei jarraituz argitaratu dira datuak: RDF estandarrean (N-Quads notazioa), tripleen datu-baseak, SPARQL 1.1 kontsulta-puntua, eta datu irekien lizentziaren erabilera (British Museums' Open Data Licence 1.1).

Ontotext teknologiaren bitartez, kontsulta-interfaz adiskidetsuagoa eskaintzen da, zeinaren bidez museoko bildumako objektuak berreskura baitaitezke haren URItik abiatuta, eta duen funtzioa gogoan izanik, hau da, bere triple storen dauzkan grafoen edo tripleteen subjektua, predikatua edo objektua den. Interfaz hau dela medio, HTMLn bistara daitezke statements direlakoak eta baita beste formatu batzuk deskargatu ere, nola diren JSON, RDF/XML, N3/Turtle, edo N-Triples.

Home | Search | Help | Licensing | About

Moaï RDF View

Source: <http://collection.britishmuseum.org/id/object/EOC3131>



Subject (53) | Predicate | Object | All

Download as: [JSON](#) | [RDF](#) | [N3/Turtle](#) | [N-Triples](#)

Named Graph: All | Language: English | Inference: Explicit only

Statements in which the resource exists as a subject.

Predicate	Object
rdf:type	ecrm:E22_Man-Made_Object
ecrm:P45_consists_of	thesx10325
ecrm:P51_has_former_or_curr...	http://collection.britishmuseum.org/id/person-institution/120561 http://collection.britishmuseum.org/id/person-institution/49731

- irudia. British Museum Semantic Web Collection (<http://collection.britishmuseum.org/resource?uri=http://collection.britishmuseum.org/id/object/EOC3131>)

6.3. VIAF LOD gisa: Linked Open Vocabularies

LOD-GLAM erresuman, VIAF (Virtual International Authority File) izeneko nazioarteko autoritate-fitxategia da zeharkako baliorik handienetakoa duen ekimenetako bat. Proiektu hau 1998an jarri zen abian, eta 2003an sendotu zen Library of Congress, Deutsche Nationalbibliothek (DNB), eta OCLC erakundeek VIAF Partzuergoa sortu zutelarik; 2007an, gainera, Bibliothèque Nationale de France (BNF) elkartu zitzairen aurrekoei. 2012az gero OCLCk eskaintzen du zerbitzu hau, eta 29 herrialdetako liburutegi nazionalekin eta zenbait herrialdetako beste ordezkariekin eusten zaio.

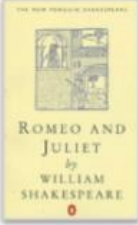
VIAFek datu-multzo bakarrean konbinatzen ditu autoritate-fitxategi ugari, iturri desberdinetakoak, iturri bakoitzeko era askotako autoritateak estekatuz (pertsonak, autoritate geografikoak, erakundeak, obrak, adierazmoldeak). Frogatu du oso aplikagarria dela, liburutegi-katalogo askotan baliatzen baita, izan Web Semantikoaren parte edo ez.

SPARQL kontsulta-punturik eskaintzen ez badu ere, VIAFi esker datu-multzoak osorik deskarga daitezke, hainbat formatotan: RDF/XML, RDF N-Triples, MARC21 XML. Irekitasunari dagokionez, Open Data Commons (ODC Attribution License) lizentziaren baitan daude datuak argitaratuak, eta horri esker merkataritza-xedeekin ere berrerabil daitezke. Baina, horrez gainera, kontsultako interfaza LOD-GLAM erresuman diren errazenetako bat da, eta, aldi berean, adiskidetsuenetako bat, eta bilaketa eta nabigazioa hainbat indiziren eta baliabide bisualen bidez egin daitezke.

VIAF
Fichero de Autoridades Virtual Internacional


Búsqueda

Selección campo: Todos los encabezamientos | Selección índice: Todo VIAF | Términos de búsqueda: baroja **Búsqueda**


 Shakespeare, William, 1564-1616
 Shakespeare, William
 شكسپير, ويليام
 1564-1616, ויליאם, שיקספיר
 Шекспир, Уильям, 1564-1616
 1616-1564, شكسپير, ويليام
 Shakespeare, William <pseud.>
 Shakespeare, William (English playwright and poet, 1564-1616)
 ID de VIAF: 96994048 (Autor personal)
 Enlace permanente: <http://viaf.org/viaf/96994048>

Formas preferidas

- 100 1 _1a Shakespeare, William, 1564-1616
- 100 1 _1a Shakespeare, William, 1564-1616
- 100 1 _1a Shakespeare, William
- 100 1 _1a Shakespeare, William, 1564-1616
- 100 1 _1a Shakespeare, William, 1564-1616
- 100 1 _1a Shakespeare, William, 1564-1616
- 100 1 _1a Shakespeare, William, 1564-1616
- 100 1 _1a Shakespeare, William
- 200 _1 1a Shakespeare, 2b William, 1564-1616
- 100 1 _1a Shakespeare, William, 1564-1616



6. irudia. VIAF, nazioarteko autoritate-fitxategi birtuala (<http://viaf.org/viaf/96994048>)

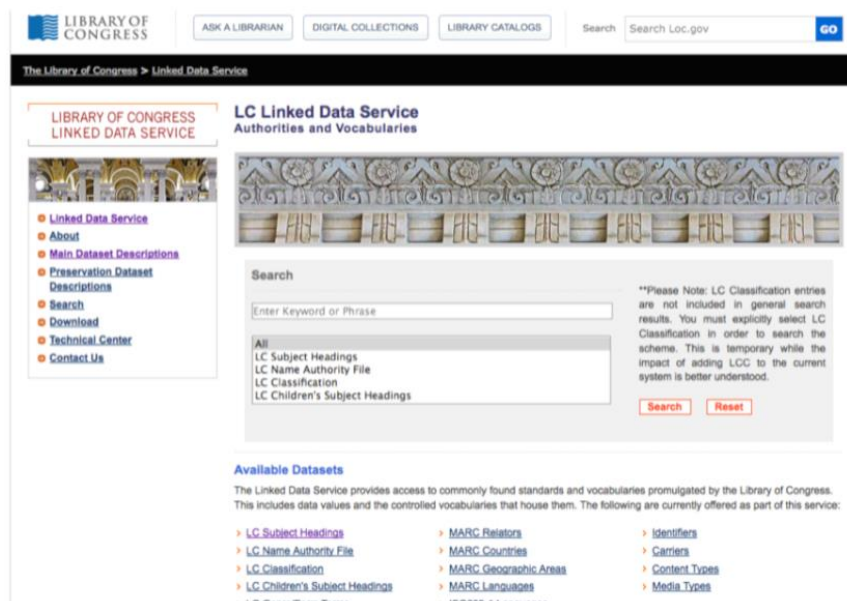
6.4. Library of Congress: Linked Data zerbitzua

Liburutegietako datuen argitalpenari dagokionez, katalogoa osorik sarean jarri gabe, eredarria da Library of Congress-eko datu estekatuen zerbitzua. Kasu honetan, autoritate-fitxategiei dagozkien zenbait datu-multzo eskaintzen dira (beren aldetik VIAFen txertatuak) eta baita beste hiztegi batzuk ere, balio kontrolatukoak, LoCen bilduma digitaletan erabiliak: sailkapenak, materia-goiburuak, tesaurusak, herrialdeen, hizkuntzen, eremu geografikoen zerrenda normalizatuak, ISO estandarrak hizkuntzak kodetzeko eta hiztegi bereziak preserbazioko metadatuentzat.

Scheme tankerako hiztegiak direlarik, datu-multzo hauek serializazio-estandar bereziak erabiltzen dituzte, nola diren SKOS/RDF (Simple Knowledge Organization System RDFn) eta MADS/RDF (Metadata Authority Description Schema RDFn), zeinaren datu-ereduaren mapa SKOSetik egin baita.

Zerbitzu honen web interfaza dela medio, aldi berean kontsulta daitezke hiztegi guztiak, edo datuak hiztegi berezien arabera iragazi, kontzeptu berreskuratuen emaitzen zerrendan barna nabigatu, termino bakoitzaren fitxa osoa testu-formatuan eta irudikapen grafikoan bistaratu, eta dagokion erregistroa hainbat kodetze eta serializazio-formatutan deskargatu (RDF/XML, N-Triples, JSON, MADS, SKOS/RDF...). Nahiz eta oraindik ez daukan bere SPARQL kontsulta-puntu propioa, datu-multzoak batera deskarga daitezke

eta berrerabili gero beste SPARQL tresna batzuetan, datu-lizentzia ireki baten erabilerari esker.



7. irudia. LC Linked Data Service: Authorities and Vocabularies (<http://id.loc.gov>)

7. Konklusio moduan: GLAMentzako aholkuak LOD erresuman

Web Semantikoa, liburutegi, museo, artxibo eta galerientzako Linked Open Data izeneko bitartez, errealitate bat dela gaur egun azaldu nahi izan dugu artikulu honetan. GLAM erakundeek LOD paradigmaren premia dute beren bildumetako datuak erakusgai, edo, hobeto, ikusgai jartzeko, eta beren metadatu berrerabilpena sustatzeko. Alabaina, aholku batzuk nahi genizkieke eman beren bildumak digitalizatu dituzten edo digitalizatzen ari diren informazioaren profesionaleri, LOD-LAMen erresuma berri honetan mugitzeko modua dela-eta:

- Erabaki datuen/metadatu zer parte diren garrantzitsuenak (Subject, Author, Publisher, etc.) LOD gisa argitaratzeko.
- Metadatuak garbitu, estekatu eta argitaratu (van Holland eta Verborgh, 2014), baina dena LOD gisa argitaratu nahi duen «mailua eskuan daukan haurra» bihurtu gabe.
- Posible denean betiere, identifikatzaile iraunkorrak erabili gauzak izendatzeko (datuak, metadatu erregistroen elementuak, hiztegien hitzak... berrerabil daitezkeen elementu posible guztiak).
- Datuak sarean/sarearentzat argitaratu (RDF). Hau da, ahalik eta triple gehienak sortu datuetatik abiatuta, datu ireki gisa argitaratu ahal izateko gero, behar izanez gero esteka litezkeenak.
- Soiltasunari eutsi: soluziorik errazena aski izan daitezke emaitza on bat izateko.

Bibliografia

*OHARRA: 2014ko abenduaren 10ean berrikusi dira azkeneko aldiz argitalpen elektroniko guztiak. Arte Garaikideko Dokumentazio Zentroen VII Topaketak

BERNERS-LEE, Tim (1999): *Weaving the Web: The Original Design and Ultimate Destiny of the World Wide Web by its Inventor*. San Francisco: Harper.

BERNERS-LEE, Tim (2006): «Linked data». *Design Issues*, 2006-07-27. Hemen eskuragarri: <http://www.w3.org/DesignIssues/LinkedData.html>

BERNERS-LEE, Tim (2009): «The next Web of open, linked data». TED, February. Hemen eskuragarri: https://www.ted.com/talks/tim_berniers_lee_on_the_next_web.html

BYRNE, Gillian and Lisa Goddard (2010): «The Strongest Link: Libraries and Linked Data». *DLib Magazine*, vol. 16, n. 11/12. Hemen eskuragarri: <http://www.dlib.org/dlib/november10/byrne/11byrne.html>

COMISION EUROPEA (2011): *Propuesta de Directiva del Parlamento y del Consejo por la que se modifica la Directiva 2003/98/CE relativa a la reutilización del sector público*. Cod. 2011/0430. [Brusela]: Europar Batzordea, 2011. Hemen eskuragarri: http://ec.europa.eu/information_society/policy/psi/docs/pdfs/directive_proposal/2012/es.pdf

COMISION EUROPEA (2014): *Comunicación de la Comisión al Parlamento europeo, al Consejo, al Comité Económico y Social Europeo y al Comité de las regiones. Hacia una economía de los datos próspera*. COM(2014) 442 final. [Brusela]: Europar Batzordea, 2014. Hemen eskuragarri: http://datos.gob.es/sites/default/files/com_2014_442_f1_communication_from_commission_to_inst_es_v2_p1_773238_4.pdf

GRADMANN, Stefan (2012): «From Containers to Content to Context: the Changing Role of Libraries in eScience and eScholarship». Bielefeld Conference, April 24 2012. Hemen eskuragarri: http://conference.ub.uni-bielefeld.de/programme/presentations/Gradmann_BC2012.pdf

HEATH, Tom and Christian Bizer (2011): *Linked Data: Evolving the Web into a Global Data Space*. Morgan & Claypool. Hemen eskuragarri: <http://www.morganclaypool.com/doi/abs/10.2200/S00334ED1V01Y201102WBE001>

HEATH, Tom, and Christian Bizer (2011): *Linked Data: Evolving the Web into a Global Data Space*. Morgan & Claypool. Hemen eskuragarri: <http://linkeddatabook.com/editions/1.0/>

MENDEZ, Eva (1999): «RDF un modelo de metadatos flexible para las bibliotecas del próximo milenio». *Jornadas Catalanas de Documentación* (1999. 7a. Bartzelona). Hemen eskuragarri: <http://www.cobdc.org/jornades/7JCD/1.pdf>

MENDEZ, EVA (2004): «La Web Semántica: una web más bibliotecaria». CLIP: Boletín de Sedic, n. 41, 2004. Hemen eskuragarri: http://www.sedic.es/p_boletinclip41_confirma.htm

MENDEZ, Eva and Jane Greenberg (2012): «Datos enlazados para vocabularios abiertos: marco global de HIVE». El Profesional de la Información, 21 (3): 236-244. Hemen eskuragarri: http://www.elprofesionaldelainformacion.com/contenidos/2012/mayo/03_esp.pdf

MINISTERIO DE LA PRESIDENCIA (España) (2011): 1495/2011 Errege Dekretua, urriaren 24koa, zeinak garatzen baitu 37/2007 Legea, azaroaren 16koa, sektore publikoko informazioa berrerabiltzeari buruzkoa. Estatuko Buletin Ofiziala, 269 zkia, 2011ko azaroaren 8koa, 116296-116307 or. Hemen eskuragarri: http://www.boe.es/diario_boe/txt.php?id=BOE-A-2011-17560

PARLAMENTO Europeo; Consejo (2013): Directiva 2013/37/UE del Parlamento Europeo y del Consejo de 26 de junio de 2013 por la que se modifica la Directiva 2003/98/CE relativa a la reutilización de la información del sector público. Diario Oficial de la Unión Europea. Hemen eskuragarri: <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2013:175:0001:0008:ES:PDF>

PROYECTO APORTA (2009): Guía Aporta sobre reutilización de la información del sector público. Madrid: Secretaria de Estado de Telecomunicaciones y para la Sociedad de la Información. Hemen eskuragarri: http://www.osimga.org/export/sites/osimga/gl/files/guia_reutilizacion_3M.pdf

VAN HOLLAND, Seth, and Ruben Verborgh (2014): Linked Data for Libraries, Archives and Museums. How to clean, link and publish your metadata, London: Facet.

W3C Incubator Group (2012): Library Linked Data Incubator Group Final Report. Thomas Baker et al. Hemen eskuragarri: <http://www.w3.org/2005/Incubator/lld/XGR-lld-20111025> (Gaztelaniaz: <http://skos.um.es/Incubator/lld/XGR-lld-vocabdataset>)