

# Il Seminario italo-spagnolo di Biblioteconomia e Documentazione

Roma, 4-5 novembre 2022

a cura di

Andrea Capaccioni e Paola Castellucci

Ledizioni

Questo libro è stato pubblicato grazie al contributo della Società italiana di scienze bibliografiche e biblioteconomiche (SISBB) e con i fondi di ricerca (A. Capaccioni) del Dipartimento di Lettere - Lingue, Letterature e Civiltà antiche e moderne dell'Università di Perugia.

Unless otherwise stated, this work is released under a Attribution-ShareAlike 4.0 International (CC BY-SA 4.0), <https://creativecommons.org/licenses/by-sa/4.0/deed.it>.



2024 Ledizioni LediPublishing  
Via Boselli 10, 20136 Milano - Italy  
[www.ledizioni.it](http://www.ledizioni.it)  
[info@ledizioni.it](mailto:info@ledizioni.it)

*Il Seminario italo-spagnolo di Biblioteconomia e Documentazione. Roma, 4-5 novembre 2022*, a cura di Andrea Capaccioni e Paola Castellucci  
Prima edizione: aprile 2024

ISBN cartaceo: 9791256001149  
ISBN ePub: 9791256001156  
ISBN PDF Open Access: 9791256001163

Progetto grafico: ufficio grafico Ledizioni  
In copertina: immagine creata da Antonella Iacono

Informazioni sul catalogo e sulle ristampe dell'editore: [www.ledizioni.it](http://www.ledizioni.it)

## INDICE

Biblioteconomia e Documentazione in Italia e Spagna: un confronto che continua. Introduzione <i>Andrea Capaccioni, Paola Castellucci</i>	7
--	---

### I. BIBLIOTECHE, TECNOLOGIE E INTELLIGENZA ARTIFICIALE

Modelli e pratiche della lettura in biblioteca. Il contesto bibliografico del progetto Reading(&)Machine <i>Maurizio Vivarelli</i>	19
--	----

Intelligenza Artificiale nelle biblioteche. Stato dell'arte ed esperienze di applicazione <i>Rossana Morriello</i>	33
--	----

L'Intelligenza Artificiale ed il problema dell'agency dal punto di vista bibliografico <i>Maurizio Lana</i>	47
---	----

Inteligencia Artificial vs. noticias falsas: las bibliotecas como agente activo frente a la desinformación <i>Sara Martínez-Cardama</i>	61
---	----

### II. ATTIVITÀ FORMATIVA E DI RICERCA: MIGLIORARE I RAPPORTI TRA I DUE PAESI

Una propuesta de internacionalización para los programas de Doctorado en Biblioteconomía y Ciencias de la Documentación entre España e Italia <i>José Luis Gonzalo Sánchez-Molero</i>	79
--	----

Iniziative di rafforzamento dell'internazionalizzazione  
della didattica in un corso di laurea 87  
*Fiammetta Sabba*

El programa doctorado de Ciencias de la Documentación  
de la Universidad Complutense de Madrid 99  
*Antonio Carpallo Bautista*

Studi mediterranei e politiche di internazionalizzazione  
tra Spagna e Sardegna. Le scienze del libro  
e del documento all'Università di Cagliari 113  
*Giovanna Granata*

### III. UN CONFRONTO SUI PROGETTI DI DIGITALIZZAZIONE DEI FONDI STORICI: ESPERIENZE INNOVATIVE, QUESTIONI, NORME

I progetti di digitalizzazione dei fondi storici: temi a confronto 123  
*Alberto Salarelli, Cristóbal Urbano*

Archivo de Revistas Catalanas Antiguas (ARCA):  
un caso paradigmático entre los proyectos de digitalización  
cooperativa en los que participa la Biblioteca de Catalunya 137  
*Eugènia Serra*

Alphabetica per la didattica e la ricerca 147  
*Elisabetta Castro*

EDL: la piattaforma digitale della Biblioteca Estense  
Universitaria di Modena 159  
*Grazia Maria De Rubeis*

# I PROGETTI DI DIGITALIZZAZIONE DEI FONDI STORICI: TEMI A CONFRONTO

Alberto Salarelli\*, Cristóbal Urbano\*\*

## *Introduzione*

Il tema della digitalizzazione dei fondi storici assume oggi una notevole rilevanza nell'ambito del panorama complessivo delle biblioteche digitali. Infatti sono molti gli spunti di innovazione che emergono sul versante tecnologico (si pensi agli sviluppi del web semantico o dell'intelligenza artificiale generativa), su quello delle forme di collaborazione fra istituzioni di differente livello (nuovi modelli cooperativi ma anche il ruolo sempre più strategico degli aggregatori), su quello delle risorse umane (nuovi profili professionali ma anche partecipazione diffusa a progetti di *citizen science*). A ciò si aggiunga come, all'interno delle diverse declinazioni nazionali del piano Next Generation EU, si sia manifestata una ingente disponibilità di risorse per il patrimonio culturale, risorse che per una parte significativa (nel caso italiano, per esempio) si è deciso di impiegare in massive campagne di digitalizzazione rivolte ai fondi storici di archivi e biblioteche di livello nazionale e non.

Per questi motivi, all'interno del II Seminario italo-spagnolo di Biblioteconomia e Documentazione, abbiamo ritenuto opportuno dedicare una specifica sessione a questo tema, coinvolgendo quattro bibliotecarie per l'esposizione di altrettanti casi di studio che, a noi curatori, sono parsi particolarmente significativi della qualità (ma verrebbe da dire: della nuova visione) con cui si stanno traducendo in atto, in progetti già in grado di essere apprezzati per l'efficacia delle soluzioni adottate, molte riflessioni ampiamente discusse nella letteratura scientifica. In questo breve saggio introduttivo Cristóbal Urbano e io vogliamo riprendere per sommi capi alcuni temi chiave per poi lasciare spazio ai singoli resoconti delle colleghe intervenute, colleghe che qui cogliamo l'occasione di ringraziare per la loro cortese disponibilità<sup>1</sup>.

---

\* Alberto Salarelli, Università di Parma, alberto.salarelli@unipr.it.

\*\* Cristóbal Urbano, Universitat de Barcelona, urbano@ub.edu.

<sup>1</sup> Tra le relazioni pubblicate in questi atti non è presente quella di Elena Sánchez (Biblioteca Nacional de España) che, tuttavia, ha partecipato a questa sessione con una relazione dal titolo *Nuevos entornos para la cooperación y la reutilización del*

## *Schemi di metadati e interoperabilità*

Affermare che i metadati rappresentino un elemento chiave nello sviluppo di un progetto di digitalizzazione significa porre l'accento su un fatto scontato. Non è certo un caso che, nell'ormai lontano 1995, il primo articolo del primo fascicolo di quella che si sarebbe rivelata una delle riviste pionieristiche nel campo delle biblioteche digitali, e cioè *D-Lib Magazine*, espressione del nascente movimento statunitense della Digital Library Federation, fosse proprio dedicato ad essi, ai metadati (Weibel, 1995), e, in particolare, a uno schema proposto, discusso e formalizzato in quello stesso anno a Dublin, Ohio...

Com'è ovvio, a quasi trent'anni di distanza, le esigenze e le pratiche di gestione delle risorse bibliografiche hanno subito notevoli cambiamenti dovuti sia ad un aumento esponenziale degli oggetti digitali resi disponibili attraverso Internet, sia ad un incremento della complessità sistemica delle soluzioni adottate per la fruizione di questi contenuti. Per questi motivi, al giorno d'oggi, la scelta e l'adozione degli appropriati schemi di metadati nelle fasi di pianificazione di una qualsivoglia iniziativa di digitalizzazione diventano passaggi in grado di rivelare la visione più o meno strategicamente orientata dell'organizzazione responsabile del progetto.

Come si può facilmente intuire, la questione assume un'importanza capitale allorché i documenti da digitalizzare siano conservati in fondi storici e/o in collezioni speciali: in questi casi è necessario valutare non solo quegli aspetti gestionali che devono essere commisurati con le caratteristiche di fragilità o di pregio dei manufatti, ma anche - per l'appunto - prestare attenzione alla scelta di soluzioni di metadattazione che rendano possibile una gestione efficace del controllo di autorità, delle descrizioni formali e semantiche, dei diritti di accesso e riproduzione, delle relazioni con ambiti di dominio differenti. Il tutto declinato in una dimensione, quella della stratificazione cronologica, che caratterizza quei documenti sui quali le tracce del tempo si sono progressivamente depositate e che perciò, in virtù di questa eredità che essi portano con sé dal passato, necessitano di un livello di analiticità del tutto peculiare. Mi limiterò, in questa sede, a focalizzare alcuni punti che mi sembrano più significativi di altri rispetto alle potenzialità, per certi aspetti davvero sfidanti, che la metadattazione contemporanea pone di fronte ai bibliotecari che si occupano di fondi storici.

Un primo aspetto da prendere in considerazione riguarda una visione complessiva e generale del ruolo dei dati della ricerca nell'epoca contemporanea che, secondo una visione ormai consolidata, devono rispondere ai criteri

---

*patrimonio digital: la propuesta BNElab de la Biblioteca Nacional de España.*

FAIR (Findable, Accessible, Interoperable and Reusable)<sup>2</sup>. I metadati, sembra quasi pleonastico ricordarlo, sono innanzitutto ‘dati’, quindi sono pienamente coinvolti in questa dinamica che mira alla massima efficienza nel reperimento e utilizzo delle risorse digitali. La questione è particolarmente sentita nel mondo delle *humanities*: infatti, premesso che «digital collections are crucial inputs to research, so the application of FAIR across the research and cultural sectors has the potential to significantly improve data sharing between researchers and heritage institutions as well» (ALLEA 2020, p. 3), e che quindi c’è un interesse primario nell’ambito della ricerca verso una metadattazione al passo con i tempi, va altresì considerato il ruolo essenziale che essa può giocare anche in un’ottica di divulgazione, migliorando la visibilità dei documenti da parte di un pubblico di non specialisti. Pensiamo, ad esempio, al compito che rivestono gli aggregatori: senza queste piattaforme che si pongono idealmente come strumenti per la scoperta, rivolti quindi a diverse categorie di utenti ma innanzitutto, vedi il caso di Europeana, agli ‘appassionati’, non potremmo fruire dei vantaggi che offrono le soluzioni *one stop* ovvero, in prima battuta, la trasversalità della ricerca in collezioni di differente natura e provenienza e, in secondo luogo, la possibilità di sfruttare *tool* innovativi sviluppati dalla comunità che ruota attorno all’aggregatore per l’allestimento di percorsi di *digital storytelling*, per il riutilizzo degli oggetti digitali ai fini didattici, per le attività di *crowdsourcing*, e via discorrendo. Ebbene, la flessibilità di utilizzo degli oggetti digitali in un aggregatore è strettamente dipendente dagli schemi di metadati utilizzati, e non è certo un caso che due piattaforme tra le più importanti a livello globale – Europeana e Digital Public Library of America – abbiano deciso di sviluppare specifici schemi tagliati sulle loro peculiarità ma, soprattutto, sulla necessità di garantire un accesso efficace a pubblici con esigenze estremamente variegat<sup>3</sup>. Ma ragionare per categorie oggi non è più sufficiente. Infatti, se nel corso della storia della metadattazione si è fatta strada la convinzione che gli schemi di metadati siano strettamente legati a determinate situazioni di contesto – sia perché, a livello di visione, punti di vista differenti possono manifestarsi in differenti domini e quindi portare a schemi che prendono in considerazione alcuni elementi e non altri, sia perché le informazioni di contesto relative alla definizione di uno schema e alle condizioni del suo utilizzo sono essenziali per ciò che concerne la

---

<sup>2</sup> Cfr. Wilkinson *et al.* (2016) e, più recentemente, Tasovac *et al.* (2020).

<sup>3</sup> Ci riferiamo all’Europeana Data Model (EDM) e al DPLA Metadata Application Profile (MAP). Come scrive Jeffrey Pomerantz (2015, p. 189): «Europeana and DPLA have taken metadata schemas created for several different use cases, selected those elements that are relevant for describing the universe of cultural heritage, and built a custom data model and element set for that purpose».

preservazione documentale e il riutilizzo dei dati<sup>4</sup> – e che quindi, in buona sostanza, i metadati non siano da considerare come entità statiche (Lorenzini *et al.*, 2020), allo stato attuale dei fatti questa concezione sembra tradursi in una tendenza sempre più spiccata verso servizi *user-oriented*, personalizzabili in base alle esigenze non di una categoria di utilizzatori ma a livello del singolo utente, in relazione da un lato agli interessi specifici di cui egli è portatore e, dall'altro, alle sue modalità concrete di accesso alla Rete, per esempio relativamente ai dispositivi di navigazione<sup>5</sup>. Se è vero che «In a scenario of digital content delivery for the cultural heritage sector – either online or onsite – to adjust what is presented to the visitor is seen as essential to accommodate different visit motivations, expectations, and needs» (Not & Petrelli, 2018), ciò vale a maggior ragione allorché le risorse digitali appartengono a un fondo storico per le ragioni a cui prima accennavamo, ragioni che si sostanziano nella necessità di dare conto di una 'durata' che il documento trattiene in sé come testimone della sua epoca e, di conseguenza, delle opportune soluzioni da adottare al fine di rendere percepibile all'utente questa dimensionalità storica. Giusto per fare un esempio, pensiamo al ruolo importantissimo che possono rivestire i metadati bibliografici contenenti informazioni sulla provenienza o sulla storia della custodia del documento: essi, come ha ampiamente dimostrato il progetto MEI<sup>6</sup>, consentono non solo di registrare informazioni cruciali per la corretta conservazione del documento stesso, ma ampliano i significati di cui esso è portatore, individuando una sua specifica collocazione nella trama dei commerci o delle correnti culturali del suo tempo (Svoljšak, 2022).

Un secondo punto oggi strategico in tema di metadattazione concerne l'interoperabilità delle soluzioni adottate. Come abbiamo detto, ampliare gli schemi di metadati, renderli più flessibili e personalizzabili è senz'altro importante nell'economia di una biblioteca digitale, tuttavia i benefici che possono derivare da questo approccio verso un'indicizzazione arricchita sarebbero fortemente penalizzati se, contestualmente, non si ragionasse in merito alla capacità delle soluzioni adottate di risultare efficaci anche nei confronti di un utilizzo esterno all'ambiente di creazione e mantenimento dei metadati stessi. La possibilità di esporre i propri metadati in un contesto di *framework* aperti è oggi cruciale per migliorare la visibilità dell'istituzione:

---

<sup>4</sup> «For data to be reusable, they must be accompanied by sufficient information to communicate the contents of the dataset, the purposes or circumstances under which it was created, and the ways in which it could be reused» (ALLEA, 2020, p. 17).

<sup>5</sup> Cfr. Konstantinos & Caridakis, 2022. Sotto questo aspetto, tornando a Europeana, è significativo che lo schema di metadati attualmente utilizzato sia stato ritenuto poco funzionale in termini di usabilità in un caso di studio condotto presso la Deutsche Digitale Bibliothek (DDB) e che, come miglioria, sia stato proposto un nuovo schema basato su FRBR (Tan *et al.*, 2021).

<sup>6</sup> <<https://www.cerl.org/resources/mei/main>>.



«The existence of many individual digital libraries, maintained by different organizations, brings challenges to the discoverability and usage of the resources by potentially interested users» (Freire *et al.*, 2020, p. 19), di conseguenza, come accennato sopra, il ruolo degli aggregatori si rivela strategico per poter intercettare l'interesse di questa utenza potenzialmente interessata ai documenti messi a disposizione da una specifica realtà. Oltre a ciò bisogna considerare anche un'altra ricaduta del concetto di interoperabilità modernamente inteso, forse ancora più importante di quella menzionata, vale a dire la possibilità di contribuire fattivamente a una crescita dell'ecosistema globale dei beni culturali: aggregando le proprie informazioni a quelle di altre istituzioni si creano le condizioni per produrre nuove informazioni di livello superiore. La questione è esposta in termini molto chiari nel *IFLA/UNESCO Manifesto for Digital Libraries*: «Interoperability and sustainability are key to the vision of digital libraries able to communicate with each other. Digital libraries that conform to commonly agreed open standards and protocols improve world-wide knowledge dissemination and access»<sup>7</sup>. Da questo punto di vista, l'adozione di soluzioni che consentano ai metadati di una specifica istituzione di essere non solo funzionali ai fini del recupero da parte di un aggregatore ma anche riutilizzabili nella più ampia visione del web semantico si rivela un passaggio strategico, proprio in considerazione del valore aggiunto che i *linked data* sono in grado di apportare alle funzionalità di ricerca e, di conseguenza, alle ricadute euristiche che da tali funzionalità derivano (Davis & Heravi, 2021). Per quanto attiene al concetto di *sustainability* richiamato nel *Manifesto*, esso può assumere differenti accezioni, dagli aspetti di natura economica legati alla possibilità di garantire alle risorse tutte quelle attività necessarie al loro aggiornamento nel corso del tempo<sup>8</sup> fino al calcolo dell'impronta di carbonio per stimare la quantità dei gas serra emessi in atmosfera dal funzionamento della biblioteca digitale sia sul lato server che su quello client<sup>9</sup>. In particolare, per ciò che attiene ai metadati, la sostenibilità si riferisce alla possibilità di garantire la loro utilizzabilità e la loro significatività nel corso del tempo, a fronte di un'obsolescenza che potrebbe verificarsi sia sul piano tecnologico che su quello semantico. Ed è per questo motivo che tale attributo si rivela strategicamente connesso all'interoperabilità: in altre parole, la scelta di schemi di metadati aperti, ampiamente diffusi nelle comunità di riferimento, supportati da una dettagliata e abbondante documentazione si pone come un passaggio fondamentale per fare in modo che le risorse possano

---

<sup>7</sup> <<https://www.ifla.org/publications/ifla-unesco-manifesto-for-digital-libraries/>>.

<sup>8</sup> Tra i primi contributi attorno a questo tema si veda Hamilton, 2004.

<sup>9</sup> Cfr. su questo aspetto in particolare i lavori di Gobinda Chowdhury, tra cui spicca Chowdhury, 2014.

essere recuperate e fruite anche domani in modo adeguato.

La terza questione in tema di buone pratiche di metadattazione ha a che fare con il fattore umano. Com'è ovvio, non è sufficiente rifarsi a soluzioni standard nell'allestimento delle collezioni digitali se poi chi deve concretamente applicarle non è in grado di farlo con le dovute cautele. Nell'ambito della tutela dei fondi storici il ruolo del catalogatore, spesso non disgiunto da quello del conservatore, ha ulteriormente evidenziato l'importanza del professionista che non solo sa interpretare le regole per la corretta descrizione della risorsa documentale ma che sa applicarle in modo consapevole perché conosce approfonditamente le caratteristiche dei manufatti posti sotto la sua responsabilità, una conoscenza che si estende tanto sul piano della loro condizione fisica, quanto su quello dei significati di cui essi sono portatori. Se pensiamo all'evoluzione delle biblioteche digitali in riferimento al ruolo di chi indicizza i documenti, il passaggio più significativo a cui stiamo assistendo è quello che richiede con sempre maggiore frequenza al catalogatore professionista l'assunzione di quelle conoscenze proprie di un *data curator*: le competenze squisitamente indicali si affiancano a quelle relative all'arricchimento dei dati da fonti esterne e alle modalità per rendere i dati riutilizzabili per esempio, come si diceva, nel contesto del web semantico. L'aspetto nevralgico di questo processo è rappresentato dalle modalità di acquisto di queste competenze e cioè dal tema della formazione dei catalogatori nell'era della metadattazione, argomento che, stante la sua complessità, non possiamo affrontare in queste poche righe; vi invitiamo a riflettere, ad esempio, attorno alle casistiche estremamente diversificate che, limitandoci all'Italia, potrebbero essere prese in esame in relazione ai corsi di laurea magistrale nell'ambito dei beni culturali: quanti di essi si rivelano sensibili a questa nuova prospettiva catalografica<sup>10</sup>? Per sottolineare come la questione delle competenze del *metadata librarian* stia vivendo una fase davvero cruciale, si aggiunga che essa deve oggi tenere in considerazione due sfide inquadrabili in prospettive fra loro alquanto differenti: da un lato, infatti, è necessario prendere in considerazione come i metadati del patrimonio culturale possano essere generati anche da utenti non specialisti come nel caso degli approcci wiki o, più in generale, dei progetti di *citizen science*. Fondamentale, in questi casi, risulta la capacità del professionista di mettere in campo tutte quelle soluzioni in grado di far rendere al massimo il potenziale informativo proveniente da tale approccio collaborativo alla metadattazione (Daga et al., 2022). Questa professionalità deve confrontarsi anche con l'altra sfida: le potenzialità sempre più performanti dei programmi di intelligenza artificiale generativa stanno già ora fornendo nuove ed amplissime prospettive alla metadattazione automaticamente assegnata, settore un tempo limitato solo

---

<sup>10</sup> Alcuni interessanti spunti di approfondimento su questo tema si ritrovano nel dibattito a più voci riportato in Guerrini, 2022, pp. 222-240.

ad alcune tipologie di metadati tecnici che invece ora si sta sempre più aprendo a sviluppi di estremo interesse verso l'indicizzazione formale e semantica (Mingfang *et al.*, 2023).

### *Las métricas de utilización: usos y usuarios en la evaluación de proyectos digitales patrimoniales*

La razón de ser de las iniciativas digitales patrimoniales se fundamenta en dos objetivos principales. Por una parte, facilitar el acceso de investigadores, y en buena medida también de público en general, a colecciones importantes y singulares del patrimonio histórico; por otra, asegurar una preservación a largo plazo de dichos materiales. Siendo sin duda importante el tema de la preservación, en términos de justificación de las inversiones, en este tipo de iniciativas resulta fundamental el objetivo de conectar con una audiencia para invitarla a descubrir contenidos con una narrativa expositiva de extensión cultural, o para satisfacer sus necesidades académicas de acceso consciente a documentos concretos por razones de investigación y estudio.

Si nos centramos en esta conexión con los usuarios, y atendiendo a lo que hemos visto en la sección anterior, cabe destacar que gracias al uso de esquemas de metadatos adecuados, la indización a texto completo en muchos casos y la interoperabilidad de sistemas, se pueden crear interfaces de usuario potentes para la recuperación de los objetos digitales en el sitio web del propio proyecto, así como potenciar la descubribilidad de esas colecciones desde el exterior gracias al posicionamiento de los contenidos en buscadores web y en agregadores patrimoniales de escala nacional o internacional.

Con dichas bases técnicas se podría pensar que ya estaría todo solucionado para atraer usuarios y generar usos, pero nada más lejos de la realidad. Para que un proyecto específico pueda maximizar la conexión con una audiencia indeterminada a escala Web y para que pueda justificar el servicio prestado a públicos específicos determinados como dianas del proyecto, se necesita de una evaluación continuada basada en datos de uso y de usuarios que permita tres cosas. En primer lugar, el diseño óptimo del sitio web y la mejora continuada de la experiencia del usuario basada en evidencias de uso; en segundo lugar, el establecimiento de una estrategia de posicionamiento, marketing y de difusión que dé a conocer las colecciones y que sirva para atraer visitantes al sitio web y, si procede, fidelizarlos; por último, la elaboración de informes que ayuden a una justificación basada en evidencias del retorno social y de la creación de valor de marca del proyecto, aspectos clave para la sostenibilidad económica del mismo. En relación a este último asunto, al igual que sucede con otras instituciones y servicios culturales, el análisis del impacto social y del valor económico que generan los proyectos digitales no es tarea fácil. Ahora bien, en este caso la intangibilidad del usuario y de sus transacciones a los ojos de los

gestores y de los políticos (no hay un espacio físico visitable donde encontrar personas a las que observar y con las que comentar su experiencia) hacen que el análisis del retorno de la inversión se torne especialmente evanescente y dependiente de la certificación de unas métricas emanadas del registro de visitas al web, en ocasiones presentadas de forma fría, sin un buen análisis contextual y sin complementos cualitativos que son de gran importancia.

En cualquier caso, sean únicamente cifras frías de métricas de utilización o resultados del análisis de la retroacción cualitativa de los usuarios, los estudios de uso juegan un papel de primer orden en la gestión de colecciones y proyectos digitales, aunque plantean nuevos retos y alternativas metodológicas (Kelly, 2014) en la recogida de estadísticas de utilización respecto a lo que venía siendo habitual en relación a los recursos y servicios analógicos de las bibliotecas. De ello son testimonio los cambios que la norma ISO 2789:2022 sobre estadísticas internacionales de bibliotecas (ISO, 2022) presenta respecto de la anterior versión del año 2013, cambios que son especialmente relevantes en relación a los recursos y servicios digitales, que se sustancian entre otras cosas en la incorporación de un anexo sobre la medición de los usos digitales, en el que se indican posibles datos a recoger, así como fuentes y métodos para captarlos (Arroyo-Vázquez, 2023). Se ha de destacar el papel que en la nueva edición de la norma se otorga a los sistemas de medición basados en programas de analítica web, ya sea los tipo *client-side* que registran la secuencia de páginas de un sitio web mostradas en el navegador del usuario durante una vista (si esas páginas están etiquetadas con un código de seguimiento), o los que realizan la explotación mediante programas *server-side* de los logs del servidor (registro de páginas enviadas a un navegador que las ha solicitado), tal y como se manifiesta en un fragmento clave de la sección 7.2 “Services and use” de la norma: “Usage data for the digitized collection and other non-comercial digital resources can generally be collected from a library’s or institution’s own servers or via tools such as Google Anayltics” (ISO, 2022). De la cita anterior cabe destacar dos hechos relevantes: por una parte, no se mencionan la funcionalidades estadísticas propias de los programas de gestión de los repositorios digitales desde los que se ofrecen los contenidos, lo que evidencia sus limitaciones; por otro lado, resulta sorprendente que en una norma ISO se mencione un programa comercial concreto como Google Analytics (GA) entre otros muchos posibles programas, y que se haga como sinónimo de las soluciones *client-side*. Ello viene a demostrar la elevada prevalencia del uso de GA en los sitios web bibliotecarios, pese a las dudas que plantea su utilización en el ámbito europeo por razones legales en materia de preservación de la privacidad y la protección de datos.<sup>11</sup>

---

<sup>11</sup> Como ejemplo de ello se puede citar el caso italiano, ya que la resolución 224 de 9/6/2022 del organismo Garante de la Protezione dei Dati Personali (Scorza,

Un análisis de urgencia visitando las principales bibliotecas digitales patrimoniales de Italia y España pone de manifiesto que casi en todos los casos se usan programas de analítica web *client-side* y que el dominio de GA es casi total, con la excepción de Matomo en un número testimonial de casos, pese a que se trata de un programa que, a criterio de la mayoría de analistas, ofrece mejores garantías desde el punto de vista de protección de la privacidad. En el caso italiano el uso de Matomo está vinculado en los casos observados de la iniciativa “Web Analytics Italia”<sup>12</sup> para el seguimiento estadístico de los sitios web de las administraciones públicas; este es el caso del portal *Alphabetic*<sup>13</sup>. Se trata de una solución especialmente robusta en cuanto a protección de la privacidad, en tanto se trata de una instalación de la versión gratuita de Matomo en servidores propios, tal y como sucede en los sitios web de la Comisión Europea y en el portal de patrimonio Europeana.

En cualquier caso, el uso de programas de analítica *client-side* en sitios web de acceso a colecciones patrimoniales es una realidad claramente establecida, que se explica más allá de la necesidad de contar con estadísticas en clave de gestión bibliotecaria tal y como se plantea en la norma ISO 2789:2022. En efecto, en último término una plataforma con contenidos patrimoniales en Internet es un sitio web por sí mismo, por lo que el análisis de su posicionamiento en buscadores, de los sitios web que enlazan con la plataforma y que aportan tráfico, de la justificación del retorno de las campañas de atracción de usuarios y del seguimiento de las visitas que recibe, son requisito fundamental para la optimización del sitio web, como lo serían para la mayoría de sitios web de cualquier otra naturaleza.

En este sentido hay que destacar que cuando hablamos de “análisis” en referencia a los datos de uso y usuarios de un sitio web, nos referimos a un concepto que implica pensar en la utilidad de la información captada para la mejora continua y el análisis del desempeño. Esto es, siendo cierto que lo que no se mide no se puede mejorar, la medición sin análisis, tampoco sirve de gran cosa. Y es que analizar implica interpretar los datos en un contexto y de acuerdo con la misión, objetivos y razón de ser del proyecto en el seno de una organización promotora del mismo. Implica que los “insights” obtenidos sobre lo que sucede en el sitio web ayuden a tomar decisiones. Analizar es algo

---

2022) dictaminó que Google Analytics transfería datos personales a los Estados Unidos sin la garantía exigida en el Reglamento General de Protección de Datos (RGPD) de la Unión Europea. Resoluciones similares se han dado en diversos países europeos. Si bien en España la autoridad de protección de datos no se ha pronunciado al respecto de ese tema en concreto, sí lo ha hecho con anterioridad respecto de la cesión de datos de Google a terceros.

<sup>12</sup> <<https://webanalytics.italia.it/>>.

<sup>13</sup> <<https://alphabetic.it/>>.

más que generar informes con datos estadísticos de utilización del sitio web. Aunque los datos se presenten muy bien y con una visualización eficaz, si no se acompañan de una explicación sobre lo que sucede y de la formulación de hipótesis de por qué sucede, su utilidad será limitada. El análisis web debería responder a lo que Avinash Kaushik denomina “Análítica Web 2.0” y que define como “el análisis de datos cualitativos y cuantitativos de su sitio web y de la competencia, para impulsar una mejora continua de la experiencia online que tienen tanto los clientes habituales como los potenciales y que se traduce en unos resultados esperados (online y offline).” (Kaushik, 2010). En esta definición de Kaushik cabe destacar la importancia que da tanto a los datos cualitativos como a los cuantitativos, a su complementariedad. Un enfoque también observado en la bibliografía sobre evaluación de proyectos digitales bibliotecarios (Shaw, 2016).

Las herramientas de analítica *client-side* a las que hemos hecho referencia son las que facilitan una información más granular y completa sobre un conjunto de métricas importantes en el contexto de las colecciones digitales (Bragg et al., 2015): cuántos visitantes se reciben en un periodo de tiempo definido; de qué zona geográfica proceden, en qué lengua tienen configurado su navegador, e incluso, si así lo configuramos, hasta su perfil de edad y sexo; cómo llegan al sitio web analizado, por ejemplo si lo hacen desde un buscador, otro sitio web, una red social, un mensaje de correo-e de una campaña de promoción, o entrando directamente la URL; qué hacen y cómo navegan cuando entran y con qué intensidad, e incluso la página por la que abandonan el sitio; y finalmente, nos pueden informar si hacen lo que esperamos que hagan, esto es, en qué medida se “convierten” a lo que nosotros previamente hayamos definido como objetivos de conversión en la configuración del programa de analítica. Además todos esos datos se pueden presentar atendiendo a la evolución temporal y a la comparación con periodos equivalentes anteriores, lo que nos arrojará indicadores en forma de porcentajes de crecimiento o de disminución de las métricas.

Ahora bien, no habría que dejarse deslumbrar por el alud de datos y de gráficos que programas como Google Analytics, Matomo y otros muchos, ofrecen. Son una aproximación a la realidad (muy representativa, ciertamente), por lo que su valor está más en las cifras relativas que podemos obtener comparando periodos de tiempo que en el valor “facial” de las métricas per se. Podemos caer en la falsa ilusión de la exactitud de datos si ignoramos que en muchas ocasiones no se puede saber con claridad cómo afectan el mecanismo de registro de las visitas los hábitos de autoprotección de la privacidad de los usuarios: algunos de ellos, o muchos, quizás configuran su navegador para deshabilitar las cookies, para que las borre cada vez que cierra el navegador, o para bloquear de forma explícita herramientas de tracking de visitas como Google Analytics. La fiabilidad de los datos recogidos tampoco está a salvo del

impacto que puedan tener ciertas acciones de spam, de ataques para saturar el servidor, o de la acción de robots que emulando un navegador ejecutan una cantidad de visitas y descargas totalmente anormales.

Haciendo las salvedades oportunas de lo que significa el verbo analizar y del valor aproximado de los datos registrados por los programas de analítica, se podrían destacar algunos usos que nos parecen especialmente importantes para el tipo de sitio web al que estamos haciendo referencia:

- Desarrollo de colecciones. Tanto la mayor frecuencia de acceso a determinadas piezas, como la estadística de las palabras claves que aportan mayor interacción en el sitio como resultado de las consultas en buscadores externos, o de las realizadas en el propio sistema de búsqueda interno, son una fuente valiosa de datos en la identificación de aquello que los usuarios consideran más atractivo o útil a sus necesidades. Como dice Biswas (2016), ese tipo de información puede servir en la toma de decisiones sobre qué contenidos hay que priorizar en la digitalización, o en la elaboración de propuestas narrativas de explotación de lo contenido en los repositorios. En muchos casos la entidad que gestiona ese repositorio o biblioteca digital patrimonial puede crear microsítios web dedicados a exposiciones, narrativas, objetos educativos, cuyo análisis tiene un potencial de primer orden para evaluar el impacto del trabajo de difusión ante los públicos diana previstos (Nicholas et al., 2014).
- Evaluación del impacto de los agregadores de contenidos patrimoniales. La explotación de las url referidas (referrers) en las que los usuarios hacen clic para llegar al sitio web objeto de análisis es uno de los datos de más valor que se pueden obtener de los programas de analítica web. A título de ejemplo, puede ayudar a valorar el impacto en términos de adquisición de visitas que tiene la presencia de los contenidos de una colección digital en agregadores de patrimonio (Biswas, 2016).
- Evaluación de campañas de atracción o de fidelización de usuarios. La amplitud del universo de los contenidos digitales en la Red, sean de naturaleza web o de interacción en redes sociales, hace que sea un reto captar la atención de los usuarios y fidelizarlos. Esto es, en el marco de la economía de la atención, el marketing digital y las campañas para captar usuarios son acciones necesarias para cumplir con el objetivo de conectar con la audiencia que debe tener todo sitio web, también uno dedicado a los contenidos patrimoniales. Sin embargo esa tarea tiene costes económicos y de dedicación de personal, por lo que es imprescindible evaluar qué acciones han funcionado y en qué medida se ha producido un retorno de la inversión en forma de aumento de las visitas o de su calidad, de la fidelización de usuarios o de la diversificación geográfica o tipológica de los mismos. En ese sentido, son especialmente útiles el registro de visitas que aparecen etiquetadas con identificadores de campañas, o que como

mínimo se producen a partir del momento que se desarrollaron acciones de promoción (Szajewski, 2013).

En resumen, de todo lo comentado y del desarrollo del debate de la tercera sesión del Seminario se puede concluir la necesidad de contar con evidencias para una buena gestión del web y de la colección digital, así como para transparentar ante las partes interesadas las métricas de utilización y de procedencia de los usuarios. Sin embargo, también se puede afirmar que la analítica web, entendida únicamente como registro estadístico de visitas al sitio web, ofrece una visión limitada que ha de ser complementada con estudios directos de usuarios basados en encuestas, la observación, los test de navegación, las entrevistas y un sistema permanente de atención al usuario que invite a la retroalimentación y a las sugerencias de mejora. En cualquier caso, todas esas acciones se han de tener en cuenta “if library services are to be all about engaging and bulding relationship with the user” (Biswas, 2016). Para conectar mejor con la audiencia, ya no digamos para ampliarla a nuevos nichos de usuarios, es muy importante disponer de información sobre los destinatarios reales y potenciales de todo proyecto digital.

### *Riferimenti bibliografici*

- ALLEA (2020), ‘Sustainable and FAIR Data Sharing in the Humanities: Recommendations of the ALLEA Working Group E-Humanities’, ALLEA-All European Academies, Berlin, DOI <<https://doi.org/10.7486/DRI.tq582c863>>.
- Arroyo-Vázquez, N. (2023), ‘Novedades en la norma ISO 2789:2022 sobre estadísticas internacionales de bibliotecas’, *Anuario ThinkEPI*, vol. 17, e17a07, DOI <<https://doi.org/10.3145/thinkepi.2023.e17a07>>.
- Biswas, P. & Marchesoni, J. (2016), ‘Analyzing Digital Collections Entrances: What Gets Used and Why It Matters’, *Information Technology and Libraries*, vol. 35, no. 4, pp. 19-34, DOI <<https://doi.org/10.6017/ital.v35i4.9446>>.
- Bragg, M. *et al.* (2015), *Best practices for Google Analytics in digital libraries*. Authored by the Digital Library Federation (DLF) Assessment Interest Group Web Analytics Working Group, DOI <<https://doi.org/10.17605/OSF.IO/CT8BS>>.
- Chowdhury, G. (2014), ‘Sustainability of digital libraries: a conceptual model and a research framework’, *International Journal on Digital Libraries*, vol. 14, pp. 181-195, DOI <<https://doi.org/10.1007/s00799-014-0116-0>>.
- Daga, E. *et al.* (2022), ‘Integrating Citizen Experiences in Cultural Heritage Archives: Requirements, State of the Art, and Challenges’, *Journal on Computing and Cultural Heritage*, vol. 14, no. 1, pp. 1-35, DOI <<https://doi.org/10.1145/3477599>>.



- Davis, E. & Heravi, B. (2021), 'Linked Data and Cultural Heritage: A Systematic Review of Participation, Collaboration, and Motivation', *Journal on Computing and Cultural Heritage*, vol. 14, no. 2, pp. 1-18, DOI <<https://doi.org/10.1145/3429458>>.
- Freire, N. *et al.* (2020), 'Cultural heritage metadata aggregation using web technologies: IIIF, Sitemaps and Schema.org', *International Journal on Digital Libraries*, vol. 21, pp. 19-30, DOI <<https://doi.org/10.1007/s00799-018-0259-5>>.
- Guerrini, M. (2022), *Metadatabase*. Milano: Editrice Bibliografica.
- Hamilton, V. (2004), 'Sustainability for digital libraries', *Library Review*, vol. 53, no. 8, pp. 392-395, DOI <<https://doi.org/10.1108/00242530410556210>>.
- ISO (2022), *ISO 2789:2022(E)-Information and documentation: International library statistics*, Vernier, Geneva: International Organization for Standardization.
- Kaushik, A. (2010), *Analitica web 2.0: el arte de analizar resultados y la ciencia de centrarse en el cliente*. Barcelona: Gestión 2000.
- Kelly, E. J. (2014), 'Assessment of Digitized Library and Archives Materials: A Literature Review', *Journal of Web Librarianship*, vol. 8, no. 4, pp. 384-403, DOI <<https://doi.org/10.1080/19322909.2014.954740>>.
- Konstantinos, M. & Caridakis, G. (2022), 'Context Awareness in Cultural Heritage Applications: A Survey', *Journal on Computing and Cultural Heritage*, vol. 15, no. 2, pp. 1-31, DOI <<https://doi.org/10.1145/3480953>>.
- Lorenzini, M., Rospocher, M., & Tonelli, S. (2020), 'Proposta per una valutazione automatica della completeness dei metadati nel contesto delle biblioteche digitali', *Digitalia*, vol. 15, no. 2, pp. 159-167, <<https://doi.org/10.36181/digitalia-00023>>.
- Mingfang W. *et al.* (2023), 'Automated metadata annotation: What is and is not possible with machine learning', *Data Intelligence*, vol. 5, no. 1, pp. 122-138, DOI <[https://doi.org/10.1162/dint\\_a\\_00162](https://doi.org/10.1162/dint_a_00162)>.
- Nicholas, D., & Clark, D. (2014), 'Information seeking behaviour and usage on a multi-media platform: case study Europeana', in: Chen C., Larsen R. (eds), *Library and Information Sciences: Trends and Research* (p. 57-78). Berlin: Springer, <[https://doi.org/10.1007/978-3-642-54812-3\\_6](https://doi.org/10.1007/978-3-642-54812-3_6)>.
- Not, E., Petrelli, D. (2018), 'Blending customisation, context-awareness and adaptivity for personalised tangible interaction in cultural heritage', *International Journal of Human-Computer Studies*, vol. 114, pp. 3-19, DOI <<https://doi.org/10.1016/j.ijhcs.2018.01.001>>.
- Pomerantz, J. (2015), *Metadata*. Cambridge, MA: MIT Press.
- Shaw, E.F. (2016), 'Making Digitization Count: Assessing the Value and Impact of Cultural Heritage Digitization', in: *Proc. IS&T Archiving 2016*, DOI <<https://doi.org/10.2352/issn.2168-3204.2016.1.0.197>>.
- Scorza, G. (2022), 'Guido Scorza su Google Analytics: "La soluzione deve essere politica, tra Stati Uniti e UE"-Intervista a Guido Scorza', *Garante per la Protezione dei Dati Personali. Stampa e comunicazione*, Doc-Web 9783595 (24/06/22), <<https://www.garanteprivacy.it/home/docweb/-/docweb-display/>>

docweb/9783595>.

- Svoljšak, S. (2022), 'Historical Collections and Library Catalogs: Provenance Metadata, Bibliographic Standards and Frameworks, and Catalog Functionalities', *Cataloging & Classification Quarterly*, vol. 60, no. 8, pp. 775-785, DOI <<https://doi.org/10.1080/01639374.2022.2124340>>.
- Szajewski, M. (2013), 'Using Google Analytics Data to Expand Discovery and Use of Digital Archival Content', *Practical Technology for Archives*, vol. 1, no. 1., <<https://hdl.handle.net/1813/76835>>.
- Tan, M.A. *et al.* (2021), 'DDB-EDM to FaBiO: The Case of the German Digital Library', in: *Proceedings of the 20th International Semantic Web Conference - Posters and Demos (ISWC2021-P&D)*, Aachen, Germany: RWTH Aachen, DOI <<https://doi.org/10.34657/8003>>.
- Tasovac, T., Chambers, S. & Tóth-Czifra, E. (2020), *Cultural Heritage Data from a Humanities Research Perspective: A DARIAH Position Paper*, <<https://hal.science/hal-02961317>>.
- Weibel, S. (1995), 'Metadata: The Foundations of Resource Description', *D-Lib Magazine*, vol. 1, no. 1, <<https://www.dlib.org/dlib/July95/07weibel.html>>.
- Wilkinson, M.D. *et al.* (2016), 'The FAIR Guiding Principles for scientific data management and stewardship', *Sci Data*, no. 3, article number 160018, DOI <<https://doi.org/10.1038/sdata.2016.18>>.