

# Dal digital curator al data librarian

MARIA CASSELLA

Biblioteca "Norberto Bobbio"  
Università degli studi di Torino  
maria.cassella@unito.it

---

## Le professioni del digitale si specializzano

---

### Un mondo data-centrico

Viviamo in un mondo data-centrico. Una mole enorme di dati è prodotta dagli utenti della rete in modo autonomo, frammentato, talvolta inconsapevole; dalle aziende, dagli enti locali, dalle organizzazioni, dalle università e dai centri di ricerca, in modo ugualmente frammentato, ma consapevole.

Le fortune di numerose aziende si fondano sui dati e sulla possibilità di riutilizzarli, scomporli, combinarli.

Grazie all'apertura in rete di buona parte dei dati (*open data*), aziende private, enti pubblici e governativi possono creare servizi a valore aggiunto per i cittadini.

Increasing numbers of individuals and organizations are contributing to this deluge by choosing to share their data with others, including Web-native companies such as Amazon and Yahoo!, newspapers such as The Guardian and The New York Times, public bodies such as the UK and US governments, and research initiatives within various scientific disciplines.<sup>1</sup>

I dati assumono, quindi, un ruolo centrale nella vita dei cittadini, li aiutano a prendere decisioni consapevoli. I dati prodotti e riutilizzati hanno anche un valore economico: siamo, come scrivono Heath e Bizer, nell'era della *data economy*.<sup>2</sup> Le università e i centri di ricerca producono ed

---

L'articolo riprende i temi affrontati dall'autrice nella sua relazione *Nuove professioni nell'era di Google: un quadro di competenze per il data librarian*, presentata in occasione del Convegno "Bibliotecari al tempo di Google" (Milano, Palazzo delle Stelline, 17-18 marzo 2016).

elaborano dati nello svolgimento delle loro quotidiane attività scientifiche. I dati prodotti per finalità di studio sono tra loro estremamente eterogenei: dati sperimentali, dati risultati da osservazioni, da simulazioni (per esempio: i modelli climatici, i modelli economici), dati derivati o compilati (per esempio: i dati estratti tramite *data/text mining*, i modelli 3D ecc. ecc.), dati di canone, dati statistici ecc. La loro gestione è fortemente legata alle differenze esistenti tra i contesti disciplinari e organizzativi.<sup>3</sup>

In passato i dati primari della ricerca erano, con poche eccezioni di archivi disciplinari dedicati alla conservazione dei dati,<sup>4</sup> scarsamente valorizzati: l'attenzione delle comunità scientifiche era focalizzata sugli output finali delle ricerche, sulle pubblicazioni, i dataset erano gestiti, quando erano gestiti, direttamente dalle comunità di ricerca, dai singoli ricercatori, erano poco diffusi, raramente pubblicati e condivisi in rete. L'esigenza di provvedere alla gestione e conservazione dei dati della ricerca è cresciuta in pochi anni in modo esponenziale. Tra le cause: l'accresciuta varietà dei prodotti della ricerca, la diffusione del movimento *open access*, la maturazione di un tipo di ricerca più consapevole e trasparente ma, soprattutto, la forte spinta degli enti finanziatori (Wellcome Trust, RCUK, NERC ecc. in Gran Bretagna,<sup>5</sup> Australian Research Council in Australia, National Institutes of Health in USA<sup>6</sup>) che da alcuni anni rendono obbligatoria in fase di presentazione dei progetti la redazione di Data Management Plans. Come e chi si sta occupando di gestire i dati della ricerca?

Negli Stati Uniti la gestione dei dati è stata affidata prevalentemente alle biblioteche accade-

niche e di ricerca. In Gran Bretagna e nei paesi europei, invece, l'approccio alla gestione dei dati è difforme essendo legato alle diverse culture istituzionali:

The data management space in US higher education is predominantly owned by the libraries (throughout the book there is an assumption that libraries are the “natural” home for data management at universities), whereas here in the UK it is much more dependent on individual institutional cultures and circumstances whether it is the librarians, the academics, or the administrators who take the lead.<sup>7</sup>

Ciononostante anche in Europa le biblioteche accademiche stanno giocando un ruolo attivo nella *data curation*<sup>8</sup> talvolta anche solo per colmare lacune organizzative. Si sono così aperte questioni nuove da affrontare.

### Digital curation, digital preservation e data curation: alcune preliminari questioni terminologiche

In relazione all'attività di ricerca il termine *digital curation* viene utilizzato per la prima volta nell'ottobre 2001 nel titolo del seminario “Digital Curation: digital archives, libraries and e-science seminar” organizzato a Londra e sponsorizzato dal Digital Preservation Coalition e dal British National Space Centre. Cosa è esattamente la *digital curation*?

Il Digital Curation Centre (DCC), un centro di ricerca sulla *digital curation* fondato nel 2004 in Gran Bretagna da un consorzio di università,<sup>9</sup> definisce la *digital curation* come l'attività che

mantiene e aggiunge valore ad un corpo certificato di informazioni digitali per l'uso corrente e futuro; nello specifico si tratta della gestione attiva e della valutazione dei dati durante l'intero ciclo di vita dei prodotti della ricerca.<sup>10</sup>

Una seconda definizione di *digital curation* è quella concepita dai promotori del Digital Curation Curriculum (DigCCURR), un progetto formativo dedicato alla *digital curation* condotto dalla School of Information and Library Science della University of North Carolina at Chapel Hill:

un decennio di lavoro su accesso e conservazione del digitale ha prodotto una complessa ed

emergente costellazione di strategie, di approcci tecnologici e di attività ora denominate: *digital curation*.<sup>11</sup>

Rispetto al rapporto tra *digital curation* e *digital preservation* un breve documento del Joint Information Systems Committee (JISC) sottolinea quanto segue:

la *digital curation* è fondata sul sottostante concetto di “conservazione digitale”, ma enfatizza l'opportunità di aggiungere valore e conoscenza alla risorsa digitale attraverso la sua gestione continua: la conservazione digitale rientra tra le attività di *curation*, sebbene entrambe si preoccupino di gestire le risorse digitali per garantirne nel tempo l'integrità.<sup>12</sup>

Proviamo, quindi, a riassumere:

- la definizione di *digital curation* proposta dal DCC è strettamente legata all'ambito accademico. Il termine “data” viene utilizzato in un'accezione molto ampia includendo qualsiasi risorsa digitale certificata che sia un prodotto dell'attività di ricerca.<sup>13</sup> In realtà, progressivamente, i confini semantici della *digital curation* si sono estesi: il termine si applica ormai in modo diffuso alla gestione e conservazione di qualsiasi documento/oggetto digitale che rientri nell'ambito dei beni culturali (*cultural heritage*). Un'attività trasversale, quindi, a vari domini del settore dei beni culturali con qualche possibile risvolto negativo per le professioni:

broad cross-domain collaboration do not serve us well if it means we ignore the vitally important differences between our various professional missions;<sup>14</sup>

- la *preservation* rientra tra le attività di *curation*. Il concetto di *digital curation* fonda su quello di *digital preservation*,<sup>15</sup> ma è più ampio e complesso. Va considerato in relazione con l'intero ciclo di vita di una risorsa digitale e con i percorsi di valorizzazione della stessa. La *curation* è un'opportunità per aggiungere valore e conoscenza ad una risorsa;

- nel contesto digitale quest'idea di gestione a ciclo unico proposta dalla *digital curation* appare estremamente significativa, dal momento che ogni risorsa viene creata e mantenuta in un

*continuum* informativo nel quale l'oggetto digitale e il medium finiscono per convivere e sono indissolubilmente legati.

E la relazione tra *digital curation* e *data curation*? Anche in questo caso i confini sono confusi e blandi. Di fatto i due termini vengono utilizzati in gran parte della letteratura come sinonimi. Infatti, il termine "data" può essere riferito sia alle fonti primarie, digitali o digitalizzate, sia ai dati della ricerca.<sup>16</sup> In ambito accademico, tuttavia, è diffuso l'utilizzo del termine *data curation* in riferimento alla gestione e conservazione dei dati della ricerca. Il contesto diventa, quindi, significativo ai fini dell'interpretazione: nel mondo accademico sembra prevalere un'interpretazione più restrittiva del termine, soprattutto quando si parla in modo specifico di *research data curation*.<sup>17</sup>

Affrontata, ma non del tutto risolta, la questione terminologica consideriamo ora le principali attività che rientrano nella *data curation*. Le elenca egregiamente Michael Witt in un passaggio del suo articolo:

un'infrastruttura flessibile e altamente scalabile deve essere in grado di ingerire enormi quantità di dati così come un gran numero di set di dati eterogenei. La conservazione e l'archiviazione dei dataset [...] richiede sistemi economicamente e tecnologicamente sostenibili per le attività di *curation*, come il controllo di integrità e il *reversioning* dei dati. Le raccolte di dati devono essere presentate in un contesto significativo e utile. Ci dovrebbero essere punti di accesso appropriati, con interfacce uomo - macchina. I metadati dovrebbero essere "harvestizzati" o creati al fine di descrivere i dataset e di sostenere funzioni come l'individuazione, l'utilizzo, la conservazione e la gestione. La provenienza dei dati deve essere registrata al fine di definire i processi di conservazione e di comprendere le *instantiations* dei dati. Meccanismi per la persistenza devono fornire identificatori univoci per i dati e un modo per risolverli attraverso citazioni persistenti che non mutano con l'evolversi dell'ambiente digitale. È necessario determinare e mantenere i diritti di proprietà intellettuale [...]. Sono necessarie policy per governare l'inserimento, la selezione, l'utilizzo e i differenti livelli di servizio dei dataset.<sup>18</sup>

In ambito accademico la *data curation* si con-

figura, quindi, come un'attività ancora più specialistica della *digital curation*. Per gestire i molteplici aspetti della *data curation* sono nate figure professionali altamente specializzate e, tra queste, quella del *data librarian*.

### Digital curator o data librarian? "That is the question!"

Alla fine del primo decennio del Duemila si sviluppa nei paesi anglofoni, soprattutto in Gran Bretagna sotto la spinta congiunta del JISC e del DCC, un grande dibattito sulla *digital curation* e sulle figure professionali atte a dominarne la complessità.

There is a need to identify the digital curation profession and this can be done by identifying the sets of skills and competences for digital curators.<sup>19</sup>

Madrid [2013] e Kim, Varge e Moen [2013] si occupano in modo specifico delle competenze del *digital curator*. Madrid utilizza il metodo Delphi per analizzare le risposte di 16 esperti interrogati sulle competenze del *digital curator*. Lo studio fa emergere competenze di tipo operativo e competenze di tipo manageriale, più competenze di dominio e abilità di tipo tecnologico (IT).

Kim, Varga e Moen<sup>20</sup> analizzano, invece, 173 annunci di lavoro pubblicati negli Stati Uniti tra Ottobre 2011 e Aprile 2012 per la figura del *digital curator* variamente denominata. L'87% degli annunci di lavoro proviene da biblioteche accademiche a conferma del fatto che il *digital curator*, pur essendo una figura trasversale ad archivi, biblioteche e musei, lavora prevalentemente nelle biblioteche accademiche e di ricerca grazie alla potente infrastruttura digitale gestita dalle università e dai centri di ricerca. Lo studio elenca sette categorie che possono raggruppare le competenze e/o abilità del *digital curator*:

- abilità interpersonali e comunicative;
- competenze relative alla gestione e conservazione dei contenuti digitali;
- competenze relative alle tecnologie;
- competenze di dominio;
- competenze di management, pianificazione e valutazione;
- competenze relative alla creazione e gestione dei servizi;

- competenze relative ai sistemi di modellazione. Fa una panoramica storica sulla figura del *digital curator* lo studio di Helen Tibbo. L'autrice analizza il percorso della formazione nell'ambito della *digital curation* negli Stati Uniti. Prende in esame i contenuti, la durata e l'audience delle iniziative formative nel Nord America, dove l'offerta formativa universitaria relativa alla *digital curation* è in lenta ma costante crescita.<sup>21</sup> In Europa il ruolo ed il percorso formativo del *digital curator* sono stati esplorati dal progetto europeo DigCurV.<sup>22</sup> Finanziato nel 2011 nell'ambito del programma Leonardo per la durata complessiva di trenta mesi, DigCurV ha costruito un quadro concettuale di riferimento per i curricula formativi del *digital curator* in Europa. Dal *digital curator* al *data librarian* il passaggio è breve: quest'ultimo è un *digital curator* che si occupa nello specifico della gestione dei dati della ricerca lavorando, come vedremo in seguito, in collaborazione con altre figure professionali. La figura del *data librarian* comincia ad emergere con chiarezza nel dibattito internazionale alla fine degli anni Duemila. Nel 2008 Alma Swan e Sheridan Brown, in un report realizzato

per il JISC, denunciano l'eterogeneità dei ruoli legati al mondo dei dati della ricerca, inserendo la figura professionale del *data librarian* tra quelle previste per la *data curation*: *data creator*, *data scientist*, *data manager* e, da ultimo ma non per ultimo, il *data librarian*.

Il report è uno dei primi documenti nel quale viene discussa la figura del *data librarian*. I due autori lo definiscono come colui che "trae origine dalla comunità dei bibliotecari, è istruito e specializzato nella 'cura', conservazione ed archiviazione dei dati".<sup>23</sup>

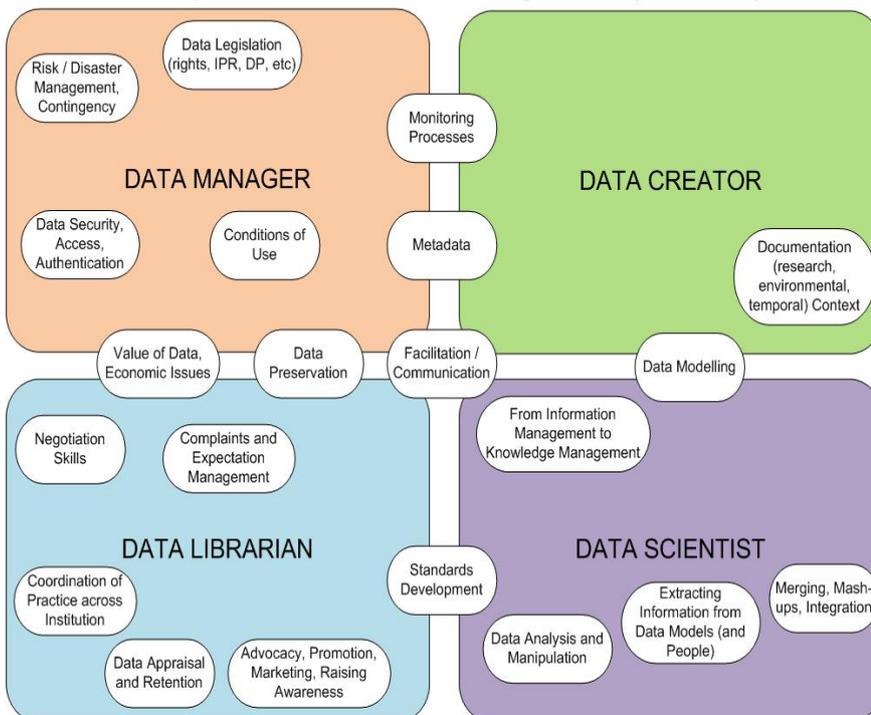
Sempre nel 2008 il Digital Curation Centre elabora un modello di competenze per la gestione dei dati della ricerca (Core skills for data management), prendendo in esame le quattro figure professionali proposte da Swan e Brown nel loro studio. Per la figura del *data librarian* vengono prese in considerazione le seguenti competenze e/o abilità:

- abilità comunicative e di coordinamento;
- sviluppo di standard;
- selezione e valutazione dei dati;
- abilità negoziali;
- advocacy, promozione e marketing;
- aspetti relativi al valore economico dei dati;
- metodologie per la conservazione dei dati;
- gestione delle aspettative e dei reclami.

In relazione al quadro di competenze proposto dal Digital Curation Centre emergono punti di forza e di debolezza. Tra i punti di debolezza va evidenziata la singolare scelta di collocare i metadati tra le competenze del *data manager* e del *data creator*,<sup>24</sup> mentre lo sviluppo di standard è presentato come un'abilità del *data librarian* e del *data scientist*. Non c'è alcun riferimento specifico alle competenze di dominio che, invece, secondo chi scrive, caratterizzano la figura del *data librarian* - insieme, ovviamente, a quella del *data scientist* - e manca un

## CORE SKILLS FOR DATA MANAGEMENT

A follow-up from the second DCC Research Data Management Forum (November 2008)



riferimento al ruolo formativo che il *data librarian* è chiamato ad esercitare in svariati contesti.<sup>25</sup> Uno studio interessante e molto recente sulla figura del *data librarian* è quello pubblicato da Brown, Wolski e Richardson [2015].<sup>26</sup> Gli autori analizzano il caso di una bibliotecaria accademica che presso la Griffith University riconfigura il proprio profilo professionale per collaborare ad alcuni progetti di ricerca della propria università. Lo studio di caso mette in luce alcuni elementi interessanti:

- l'importanza strategica delle competenze legate al settore disciplinare;
- la necessità di acquisire una discreta conoscenza dei database relazionali e dei *content management systems*;
- la necessità di possedere una discreta conoscenza ed esperienza in materia di schemi di metadati. Gli autori propongono, quindi, la seguente griglia di competenze per il *data librarian* che deve possedere:
- una profonda comprensione dei processi e dei risultati della ricerca, delle modalità della comu-

nicazione scientifica, delle tipologie e dei formati di dataset secondo le specificità disciplinari;

- una profonda comprensione delle questioni relative all'etica professionale, ai diritti di proprietà intellettuale, alle licenze, al copyright;
- una profonda comprensione delle attività di storage, backup e trasferimento dei dati;
- un'ottima conoscenza degli schemi di metadati e dei relativi standard, una comprensione dei linguaggi di marcatura XML;
- una buona conoscenza degli schemi e degli standard di certificazione dei repository.
- una buona conoscenza degli standard del web semantico e delle piattaforme per gli open data e i linked open data;
- abilità e/o competenze elevate relativamente a: comunicazione, *project management*, *system design*, *business analysis*.

Proviamo, quindi, a tirare le fila di quanto abbiamo scritto sopra e a proporre di seguito un nostro quadro di competenze per il *data librarian*, individuando cinque principali macroaree di riferimento:

Biblioteconomia	Le competenze biblioteconomiche sono di base. Il <i>data librarian</i> possiede una formazione da umanista. È specializzato negli schemi di metadati <sup>I</sup> e nella conoscenza dei relativi standard. Questa competenza caratterizza la sua figura professionale differenziandola da quelle del <i>data scientist</i> e del <i>data manager</i> .
Comunicazione scientifica e ricerca	Il <i>data librarian</i> è un esperto di comunicazione scientifica. È informato sui bandi di ricerca. Conosce le policy degli enti finanziatori; è esperto dei processi di ricerca in relazione ad uno specifico dominio. Una buona conoscenza del settore disciplinare è un valore aggiunto per il <i>data librarian</i> .
Tecnologia	Schemi e standard per la conservazione e per la certificazione dei repository, linguaggi di marcatura (XML) e, ancora, sempre più importanti, gli standard del web semantico, RDF data model, i linguaggi per la costruzione di ontologie, le piattaforme per la pubblicazione dei dataset, <sup>II</sup> i linguaggi di visualizzazione e virtualizzazione ecc. <sup>III</sup>
Diritto. Diritti e licenze	Un must nel mondo digitale. Alcune licenze di uso sono state concepite in modo specifico per i dataset e i database. <sup>IV</sup> Dalla versione 4.0 le licenze Creative Commons possono essere utilizzate per i dataset.
Management e Comunicazione	Abilità comunicative, attitudine al cambiamento e al lavoro di gruppo, competenze in relazione al project management. Il <i>data librarian</i> può anche essere un educatore, un formatore, coadiuva i ricercatori nella redazione dei Data Management Plans.

I. Segnalo che sul sito del Digital Curation Centre è disponibile un elenco di schemi di metadati esistenti per i diversi ambiti disciplinari. Alla URL: <http://www.dcc.ac.uk/resources/metadata-standards>.

II. Ad esempio: TheDatahub. <https://datahub.io/>

III. Ovviamente il livello di approfondimento di questi argomenti è molto diverso tra un bibliotecario e un informatico.

IV. Le licenze Open Data Commons e la IODL 2.0, la licenza italiana.

Il quadro di competenze proposto è generale e va adattato alle attitudini personali e all'esperienza lavorativa pregressa di chi si candida al ruolo di *data librarian*. Il suo scopo è sostenere eventuali esperienze di formazione per la figura del *data librarian*.

A questo punto ci sembra utile proporre alcune riflessioni di sintesi prima di avviarcì alle conclusioni.

1. La *data curation* è un'attività alla quale sono legati una varietà di processi e una molteplicità di figure professionali e di *stakeholders*. Non può essere gestita da un'unica figura professionale. Il *data librarian* lavora in staff, deve essere abile nel comunicare e nel sviluppare nuove alleanze, interne ed esterne: con i ricercatori, con i manager della ricerca, con i centri di ricerca, con gli enti finanziatori ecc.

2. Il *data librarian* lavora a stretto contatto con le comunità di ricerca, sovente partecipa ai progetti di ricerca e, quindi, dal punto di vista organizzativo lavora nei dipartimenti. In alternativa è il repository manager che può assumere anche il ruolo di *data librarian* con compiti maggiormente legati alla gestione e conservazione dei dataset nei repository. In questo caso – e solo in questo caso – le competenze tecnologiche finiranno per essere prevalenti.

3. La molteplicità e varietà di ruoli bibliotecari a partire dai quali può formarsi e gemmare la figura del *data librarian* è enorme: dal *subject librarian*, ai bibliotecari addetti ai servizi di *reference* e, più in generale, ai servizi al pubblico, ai bibliotecari che si occupano di ricerca e di comunicazione scientifica, fino ai repository manager, ai *system librarian* e agli esperti di schemi di metadati.

## Conclusioni

Tra le professioni nate con e per la biblioteca digitale quella del *data librarian* è certamente tra le più complesse.

I suoi confini sono ancora sfumati:

given the variety of roles discussed and the myriad of skill sets needed to perform these roles, it is not terribly surprising that e-science and data librarianship has been described as nebulous and undefined.<sup>31</sup>

Negli assetti organizzativi dei sistemi di biblioteca e/o delle biblioteche si tratta di un ruolo ancora poco presente e scarsamente riconosciuto. Sovente il bibliotecario accademico si occupa dei dati della ricerca senza avere un ruolo definito.

Con la figura del *data librarian* si innalza il livello di specializzazione del *digital curator*. Questa iper-specializzazione porta con sé alcune inevitabili conseguenze, in primo luogo la necessità di una formazione adeguata.

Come si forma il *data librarian*? Quali sono le sue opportunità formative? Al momento la formazione universitaria per questo tipo di figura professionale è quasi assente, mentre esistono numerose esperienze formative curriculari per il *digital curator*,<sup>32</sup> la figura del *data librarian* non sembra ancora essere stata presa in considerazione nei curricula accademici, sia perché di recente costituzione e dai confini ancora indefiniti, sia perché il ruolo del *data librarian* tende a configurarsi come una specializzazione del *digital curator*.

Prevalgono, al momento, le opportunità di auto-formazione rivolte a quanti già lavorano come professionisti nel settore delle biblioteche: seminari, convegni, workshop e corsi e-learning erogati con strumenti formativi di diverso tipo (*toolkits*, MOOCs).<sup>33</sup> Un'utile sitografia sul tema dell'auto-formazione nell'ambito del *research data management* è quella di Goben e Raszewski.<sup>34</sup> Quanto allo strumento didattico del *toolkit* vale la pena segnalare il Data Curation Profiles Toolkit<sup>35</sup>, sviluppato grazie ad un progetto di ricerca delle biblioteche della Purdue University, della Graduate School of Library and Information Science e dell'Università dell'Illinois at Urbana-Champaign.

Resta, comunque, la necessità di esplorare anche percorsi curriculari più sistematici, di tipo accademico, man mano che le esigenze di gestione dei dati della ricerca cresceranno:

There is a need to go beyond the workshop and the training course, and embed preparation for a professional (and personal) lifetime of digital data curation within the academic curriculum.<sup>36</sup>

In secondo luogo va evidenziato il rischio che, in mancanza di una definizione del ruolo del *data librarian* e di un suo incardinamento negli or-

DCP Home Overview Publish a Profile DCP Directory Archive Contact [Download the Toolkit](#)

# Data Curation Profiles

Thank you for your interest and spreading the use of the [Data Curation Profiles Toolkit](#). In October 2014 we received an Institute of Museum and Library Services Planning Grant to continue working on on the DCP. The primary goal of this grant is to create a roadmap to scope the outcomes and work needed for a redesign. We believe this can contribute to curating more of research outputs—moving from active management of data and digital objects to dissemination and preservation of them. A “bridging the gap” report will be published later that pulls together the thinking of experts on issues and challenges for that. Stay tuned...

Data Curation Profiles are in transition While we are in planning stages, we are migrating content from this site to Purdue University Libraries' document repository, e-Pubs. The DCP Toolkit can still be downloaded as noted, and is the current version to date. Please contact us if you have any questions.

The [Data Curation Profiles](#) can help you:

-  Guide a discussion about data with researchers
-  Gain insight into areas of attention in data management
-  Assess information needs related to data collections
-  Create a starting point for curating and preserving data

Home page del sito [datacurationprofiles.org](http://datacurationprofiles.org)

ganigrammi delle biblioteche, il *data librarian* resti una figura occasionale, che nasce e muore con i progetti di ricerca, di sovente gemmata da altre figure professionali che non necessariamente appartengono all'area delle biblioteche: ricercatori, manager della ricerca, dottorandi ecc.

Come sempre accade, la tecnologia e l'innovazione rappresentano anche una minaccia per le professioni. Nelle biblioteche accademiche, in modo particolare, si è concretizzata più volte questa minaccia.

Per garantire un futuro alla professione i bibliotecari accademici dovranno necessariamente essere flessibili e riposizionarsi, stando attenti a non snaturare la propria identità professionale.

We must not lose sight of the distinct identity of our profession but nor should we be complacent about it. We must continue to reconsider and reinvent our profession in response to changing circumstances and needs while always remembering our higher goals.<sup>37</sup>

Abbiamo discusso del *data librarian* nel contesto della biblioteca accademica e, quindi, dell'università, dove vengono principalmente prodotti i dati della ricerca.

Purtuttavia, in rete una gran mole di dati viene creata quotidianamente per interazione degli utenti. L'analisi di questi dati serve alle aziende per migliorare la comunicazione e posizionarsi sul mercato. Di fatto, una percentuale sempre più significativa (80%) delle informazioni rilevanti in ambito aziendale è rappresentata dai dati non strutturati del web sociale.<sup>38</sup> Anche nel settore privato sono, quindi, nate nuove figure professionali, come, ad esempio, quella del *data scientist* o *data analyst*, come viene definito colui che analizza i dati sul web al fine di orientare le scelte strategiche aziendali.<sup>39</sup>

Lentamente, l'analisi dei Big Data sta crescendo di importanza anche per l'ambito delle biblioteche trasformando lo scenario della valutazione. Aiuta a comprendere le performance della

biblioteca, a sostenere le decisioni e a migliorare il servizio al pubblico.<sup>40</sup>

In questa prospettiva il *data librarian* assume una veste ancora diversa: è uno specialista dell'analisi dei dati del web, lavora per le biblioteche pubbliche e per quelle accademiche. In breve, si trasforma in una figura professionale simile al *data analyst* di un'impresa privata.

Questa declinazione in chiave aziendale apre ulteriori possibilità di sviluppo per la professione bibliotecaria. Nuove interessanti contaminazioni potrebbero emergere in futuro tra professionisti del settore pubblico e professionisti del settore privato.

---

---

## NOTE

<sup>1</sup> TOM HEATH - CHRISTIAN BIZER, *Linked data: evolving the web into a global data space*, Synthesis Lectures on the Semantic Web: Theory and Technology, London, Morgan & Claypool Publishers, 2011. <http://linkeddatabook.com/editions/1.0/#htoc2>.

<sup>2</sup> TOM HEATH - CHRISTIAN BIZER, op. cit., 2011.

<sup>3</sup> Cfr. MICHAEL WITT et al., *Constructing data curation profiles*, in "The International Journal of Digital Curation", 4(2009), n. 3. [http://docs.lib.purdue.edu/cgi/viewcontent.cgi?article=1141&context=lib\\_research](http://docs.lib.purdue.edu/cgi/viewcontent.cgi?article=1141&context=lib_research). Gli autori intervistano 19 ricercatori di diverse aree scientifiche per individuare le priorità delle comunità di ricerca e creare dei profili per la gestione dei dataset.

<sup>4</sup> Ad esempio in Gran Bretagna il UK Data Archive che ha festeggiato nel 2007 quaranta anni di attività. È un repository dedicato alle scienze sociali. Si legga: UK Data Archive, *Across the decades. 40 years of data archiving*. <http://data-archive.ac.uk/media/54761/ukda-40thanniversary.pdf>.

<sup>5</sup> <http://www.dcc.ac.uk/resources/policy-and-legal/overview-funders-data-policies>.

<sup>6</sup> Quanto ai National Institutes of Health segnalo la tabella sulle NIH Data Sharing Policies, [https://www.nlm.nih.gov/NIHbmic/nih\\_data\\_sharing\\_policies.html](https://www.nlm.nih.gov/NIHbmic/nih_data_sharing_policies.html) e, per una lettura, il documento *NIH Public Access Plan for Increasing Access to Scientific Publications and Digital Scientific Data from NIH Funded Scientific Research*, Febbraio 2015 disponibile alla URL <https://grants.nih.gov/grants/NIH-Public-Access-Plan.pdf>.

<sup>7</sup> Cfr. DAVID FEARON et al., *Research data management services*, (SPEC Kit N. 334), 2013.

<sup>8</sup> In Europa il ruolo delle biblioteche accademiche rispetto alla gestione dei dati della ricerca è stato sottolineato da organizzazioni di rilievo come la League of European Research Universities (LERU) che nel 2013 pubblicava la sua Roadmap sui dati della ricerca (LERU Roadmap for research data) e dalla Ligue des Bibliothèques Européennes de Recherche (LIBER) nel documento: "Ten Recommendations for Libraries to Get Started in Research Data Management".

<sup>9</sup> In particolare, le università di Bath, Edimburgo e Glasgow.

<sup>10</sup> "Maintain and add value to a trusted body of digital information for current and future use; specifically, (it is) the active management and appraisal of data over the life-cycle of scholarly and scientific materials". Da: DCC (Digital Curation Centre), *What is digital curation?*. <http://www.dcc.ac.uk/digital-curation/what-digital-curation>. Ultimo accesso: 10.01.2016. La traduzione è mia.

<sup>11</sup> Cfr. CHRISTOPHER A. LEE - HELEN R. TIBBO - JOHN C. SCHAEFER, *DigCCURR: building an international digital curation curriculum & the Carolina Digital Curation Fellowship Program*, in "Archiving" 2007, <http://ils.unc.edu/caltee/p105-lee.pdf>. La traduzione è mia.

<sup>12</sup> "Digital curation builds upon the underlying concepts of digital preservation whilst emphasising opportunities for added value and knowledge through annotation and continuing resource management. Preservation is a curation activity, although both are concerned with managing digital resources with no significant (or only controlled) changes over time". In JISC, *Digital preservation: continued access to authentic digital assets. Briefing paper*, November 2006. <http://www.webarchive.org.uk/wayback/archive/20140615231719/http://www.jisc.ac.uk/media/documents/publications/digitalpreservationbp.pdf>. La traduzione è mia.

<sup>13</sup> Secondo il DCC il valore complessivo della ricerca finanziata in UK è annualmente di 3 miliardi di sterline.

<sup>14</sup> ADRIAN CUNNINGHAM, *Digital curation/digital archiving: a view from the National Archives of Australia*, "American Archivist", 71(2008), n. 2, p. 530-544. La citazione è a pagina 532.

<sup>15</sup> Scrive a tal proposito in modo illuminante A. Cunningham: "Included within the definition of digital curation are the noble endeavors of digital preservation, digital librarianship, digital archiving, and digital management". Cfr. ADRIAN CUNNINGHAM, cit. 2008. La citazione è a pagina 531.

<sup>16</sup> Il termine è, comunque, polivalente. Ad esempio, nell'ambito ambito storiografico ha praticamente sostituito l'uso del sostantivo "fonte", mentre dal punto di vista sociologico i dati sono "fatti". Cfr. SERGE NOIRET, *Storia contemporanea digitale*, in "Il web e gli studi storici: guida critica all'uso delle rete", a cura di Rolando Minuti, Roma, Carocci, 2015, p. 267-300.

<sup>17</sup> MICHAEL WITT, *Institutional repositories and research data curation in a distributed environment*, in "Library Trends", 57(2008), n. 2, p. 191-201. [http://docs.lib.purdue.edu/cgi/viewcontent.cgi?article=1126&context=lib\\_research](http://docs.lib.purdue.edu/cgi/viewcontent.cgi?article=1126&context=lib_research).

<sup>18</sup> "Flexible and highly scalable infrastructures must be able to ingest massive datasets as well as large numbers of heterogeneous datasets. The preservation and archiving of data so that they can be accessed and used in the distant future necessitates economically and technologically sustainable systems for ongoing curation activities such as data integrity checking and reversioning. Data collections need to be presented in a meaningful and useful context. There should be appropriate points of access with both human and machine interfaces. Proper metadata must be captured or created to describe the data to support functions such as discovery, use, preservation, and administration. The provenance of the data needs to be recorded in order to establish a chain of custody and understand the instantiations of

the data. Mechanisms for persistence are required to provide unique identifiers for data and a way to resolve them from citations that will not break as the information environment evolves and changes. Intellectual property rights must be determined and maintained, and permissions for accessing the data must be enforced. Policies are needed to govern submissions, selection, usage, and levels of service, at a minimum". MICHAEL WITT, op. cit. 2008, p. 195. La traduzione è mia.

<sup>19</sup> MELODY M. MADRID, *A study of digital curator competences: a survey of experts*, in "The International Information and Library Review", 45(2013), p. 149-156.

<sup>20</sup> JEONGHYUN KIM - EDWARD WARGA - WILLIAM E. MOEN, *Competencies required for digital curation: an analysis of job advertisements*, in "The International Journal of Digital Curation", 8(2013), n. 1. <http://www.ijdc.net/index.php/ijdc/article/view/8.1.66>.

<sup>21</sup> HELEN R. TIBBO, *Digital curation education and training: from digitization to graduate curricula to MOOCs*, in "The International Journal of Digital Curation", 10(2015), n. 1. <http://www.ijdc.net/index.php/ijdc/article/view/10.1.144/387>.

<sup>22</sup> <http://www.digcur-education.org/>

<sup>23</sup> "The data librarian originates from the library community, is trained and specialized in the curation, preservation and archiving of data". Cfr. ALMA SWAN - SHERIDAN BROWN, *The skills, role and career structure of data scientists and curators. An assessment of current practice and future needs. Report to the Joint Information Systems Committee (JISC)*, 2008, <http://eprints.ecs.soton.ac.uk/16675>. La traduzione è mia.

<sup>24</sup> Le due principali figure collegate con la creazione e gestione dei metadati sono, invece, a mio avviso: il *data librarian* e il *data creator*. Quanto a quest'ultima figura è coinvolta nella creazione di alcuni metadati che vengono generati in automatico nel momento della creazione della risorsa digitale.

<sup>25</sup> Ad esempio, in merito a: selezione dei dati, *workflow* dei dati, licenze da adottare, metodologie e procedure di conservazione ecc. ecc.

<sup>26</sup> REBECCA A. BROWN - MALCOLM WOLSKI - JOANNA RICHARDSON, *Developing new skills for research support librarians*, in "The Australian Library Journal", 64(2015), n. 3, p. 224-234.

<sup>27</sup> Segnalo che sul sito del Digital Curation Centre è disponibile un elenco di schemi di metadati esistenti per i diversi ambiti disciplinari. <http://www.dcc.ac.uk/resources/metadatat-standards>.

<sup>28</sup> Ad esempio: TheDatahub <https://datahub.io/>.

<sup>29</sup> Ovviamente il livello di approfondimento di questi argomenti è molto diverso tra un bibliotecario e un informatico.

<sup>30</sup> Le licenze Open Data Commons e la IODL 2.0, la licenza italiana.

<sup>31</sup> JAKE R. CARLSON, *Opportunities and barriers for librarians in exploring data: observations from the data curation profile workshops*, in "Journal of eScience Librarianship", 2(2013), n. 2. <http://escholarship.umassmed.edu/jeslib/vol2/iss2/2/>

<sup>32</sup> Ad esempio, il Master internazionale DILL e il progetto formativo statunitense DigCCurr.

<sup>33</sup> Ad esempio, l'University College London ha realizzato nel 2014 un MOOC dedicato alla *data curation*.

<sup>34</sup> ABIGAIL GOBEN - REBECCA RASZEWSKI, *Research data management self-education for librarians: a webliography*, in "Issues in Science and Technology Librarianship", Fall 2015. <http://istl.org/15-fall/internet2.html>

<sup>35</sup> <http://datacurationprofiles.org/index.php>.

<sup>36</sup> Cfr. GRAHAM PRYOR - MARTIN DONNELLY, *Skilling up to data: whose role, whose responsibility, whose career?*, in "The International Journal of Digital Curation", 4(2009), n. 2, p. 166.

<sup>37</sup> ALEX BYRNE, *Libraries for the future: progress, development and partnerships. Opening address. World Library and Information Congress: 73rd IFLA Council and General Congress, Durban, South Africa*, in "IFLA Journal" 33(2007), n. 3, p. 258-260.

<sup>38</sup> Cfr. ROBERTA RAIMONDI, *Come valorizzare le fonti documentali*, in "Dai documenti al social web", a cura di Roberta Raimondi, Milano, Franco Angeli, 2012, p. 125.

<sup>39</sup> Cfr. *Le nuove professioni digitali: risorse, opportunità e competenze per la tua carriera online*, a cura di Giulio Xaet e Ginevra Fidora, Milano, Hoepli, 2015.

<sup>40</sup> Cfr. CHIARA FAGGIOLANI - MAURIZIO VIVARELLI, *Leggere in rete. La lettura in biblioteca al tempo dei big data*, in "Bibliotecari al tempo di Google: profili, competenze, formazione", Milano, Bibliografica, 2016, p. 101-126.

DOI: 10.3302/0392-8586-201603-013-1

## ABSTRACT

The data librarian is a new professional role in librarianship. Stemming from manifold different roles, the data librarian is a specialization of the digital curator but its boundaries are still nebulous and undefined. The paper analyzes the main differences between the digital curator and the data librarian, identifies the leading roles in data management, i.e. data creator, data manager, data scientist, and data librarian and explores the skills and the competencies of the latter one in the academic library environment. It proposes a core skills for data librarianship referring to five macroareas: library and information science, scholarly communication, technology, intellectual property rights, and management. In conclusion, the paper discusses some issues regarding the development of digital and data curation curricula.