

Meaningful and Inclusive Access to Information: the Challenges Brought by the Brisbane Declaration to Standardized Metadata in the Context of Linked Data and AI

Philip Schreur^(a)

a) Stanford University Libraries

Contact: Philip Schreur, pschreur@stanford.edu

Received: 15 May 2025; Accepted: 30 June 2025; First Published: 15 September 2025

ABSTRACT

With the rise of Google and Wikidata as authoritative sources of information for many, with the use of generative Artificial Intelligence, such as Chat GPT, for the creation of new content, and with linked open data as the global means of communication for the Semantic Web, the structuring and dissemination of data has become deprofessionalized. In such a shifting landscape, what is the role of new technologies, such as Artificial Intelligence (AI), in addressing key challenges to the creation of standardized metadata and its dissemination?

The launch of the Brisbane Declaration at the close of IFLA's Information Futures Summit stresses "what principles should apply when taking decisions about the future, highlighting changing knowledge practices, the impacts of new technology, [and] the importance of trust, skills, and equity." Standardized metadata has been the lingua franca of information exchange in the library community since at least the creation of the *Paris Principles*. The popularization of metadata creation, although not challenging the intent of the *Principles* themselves, forces us to reconsider their realization. And new technologies, such as AI and linked open data, give us new opportunities to explore meaningful and inclusive access to information.

KEYWORDS

Bibliographic metadata; Linked Open Data (LOD); Artificial intelligence; Brisbane Declaration.



1. Introduction

In 2017, Tom Nichols published a book entitled *The Death of Expertise: the Campaign Against Established Knowledge and Why it Matters* (Nichols 2017). In his book, Nichols details how the rise of the information age through the integration of the Internet into every aspect of our lives has created a misguided sense of intellectual egalitarianism. Everyone considers themselves to be an expert on everything; and now Artificial Intelligence (AI) is trusted above everything else.

The focus of Nichols' book is on the political arena, but it doesn't take much of a shift to see how his observations might be applied to metadata standards and their role in the development of a sustainable, equitable, and accessible future for information. With the rise of Google and Wikidata as authoritative sources of information for many, with the use of generative Artificial Intelligence, such as Chat GPT, for the creation of new content, and with linked open data as the global means of communication for the Semantic Web, the structuring and dissemination of metadata has become deprofessionalized.

Standardized metadata, however, has been the lingua franca of information exchange in the library community since at least the creation of the *Paris Principles (Statement of Principles* 1961). The popularization of metadata creation through such tools as WikiData, although not questioning the intent of the *Principles* themselves, forces us to reconsider their realization as technologies, such as AI and linked open data, give us new opportunities to explore the creation of meaningful and inclusive access to information by the general public.

In October of 2024, IFLA held an Information Futures Summit directly confronting some of these issues. One of the outcomes of this meeting was *The Brisbane Declaration* (IFLA 2024) which challenges us to focus on technological change, social justice, and barrier free access to information so that no one faces information poverty or exclusion. Unfortunately, it is not surprising to many of us that our Metadata Departments are overwhelmed with resources in unfamiliar languages, new formats, and sweeping digital collections.

And so, as libraries embrace the exploration of linked open data and Artificial Intelligence in the hopes of providing access to this avalanche of new resources, what ultimately will become the balance of trust and provenance between established metadata standards and a more populist approach to metadata creation and identity management? How do we balance the use and availability of these new tools with our focus on the expertise that professional catalogers bring to the issues of description and discovery through the exacting metadata standards they employ? In exploring this issue, we must first review how and why we currently apply our metadata standards. And then, how *The Brisbane Declaration*, in conjunction with the development of linked open data and Artificial Intelligence, could challenge us to move in completely new directions.

2. History of Metadata Standards Application

This first example may still seem familiar to many, an example of a catalog card:

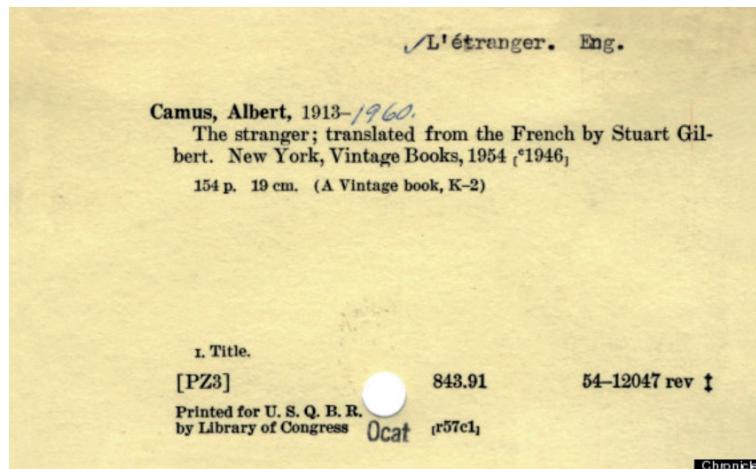


Figure 1. Albert Camus catalog card.

The principle is clear, we review the contents of a resource, in this case a book, and reduce it to a set of required metadata elements. The metadata standards in its creation are very exacting, from the name of the author to the metadata that is captured. This collection of metadata elements, the catalog record, becomes a stand-in, or surrogate, for the physical item. And because the metadata standards that are applied are consistent, we can file them with other similarly created cards to allow for discovery of all the resources in our collection.

But you can already see the downside to this technology. The death date of Albert Camus was added in pen directly to the catalog card. And the Uniform Title for this work was added by typewriter to the top of the card.

In the 1960s, we had the brilliant innovation of the MARC formats.

▼ Metadata

LEADER 00936cam a2200301 i 4500
001 a319327
003 SIRSI
005 20240311180212.5
008 840606t19541946nyu 000 1 eng d
010 a| 72008033
020 a| 0394700023
020 a| 9780394700021
035 a| (Sirs) ABQ3401
035 a| (CStRLIN)CSUG84-B39072
035 a| (OCoLC-M)1453514
035 a| (OCoLC-I)272297366
040 a| NvU c| NvU d| laU d| OrLoB d| CSt
041 1 a| eng h| fre
100 1 a| Camus, Albert, d| 1913-1960.
245 1 4 a| The stranger / c| Albert Camus ; translated from the French by Stuart Gilbert.
260 a| New York ; b| Vintage Books, c| 1954, c1946
300 a| 154 p. ; c| 18 cm.
500 a| Translation of L'étranger.
596 a| 1 20 31
700 1 a| Gilbert, Stuart.
999 f f i| a796a5f7-9044-544e-ac3f-436288606ebe s| 139a8c05-eacf-514b-8b07-b5785130ae21

Figure 2. Albert Camus MARC.

Although this was an innovation in technology, it allowed us both to update our metadata in a breathtakingly simpler manner and to share our metadata locally, nationally, and internationally. But for all its advantages in communication, we still only used the MARC formats to represent the catalog card on a computer screen.

And in May of 2011, we had another great shift in technology with the implementation of linked data in the form of BIBFRAME.

```
<https://api.stage.sinopia.io/resource/64aeb7b7-5e9d-4d77-b5c0-172d78ef74fd>
  a bf:Instance, bf:Print;
  bf:instanceOf <https://api.stage.sinopia.io/resource/d64f2339-2bba-44b4-9eb6-7399c071bbe2>;
  bf:title a bf:Title ;
    bf:mainTitle "The stranger" ;
    bf:nonSortNum "4" .
  bf:responsibilityStatement "Albert Camus ; translated from the French by Stuart Gilbert"@en;
  bf:provisionActivity a bf:Publication ;
    bf:date "1954";
    bf:place "New York (State)";
    bf:lc:simplePlace "New York"@en;
    bf:lc:simpleAgent "Vintage Books"@en;
    bf:lc:simpleDate "1954".
  bf:copyrightDate "©1946";
  bf:identifiedBy a bf:ISBN;
    rdfs:value "9780394700021", rdf:value "0394700023" .
  bf:extent a bf:Extent ;
    rdfs:label "154 pages"
  bf:dimensions "18 cm";
  bf:issuance a bf:Issuance ;
    rdfs:label "single unit" .
  bf:media a bf:Media ;
    rdfs:label "unmediated" .
  bf:carrier a bf:Carrier ;
    rdfs:label "volume" .
  bf:adminMetadata a bf:AdminMetadata;
    bf:lc:catalogerId "NancyL";
    bf:creationDate "2025-01-06";
    bf:assigner a bf:Agent ;
      rdfs:label "Stanford University" .
    bf:lc:encodingLevel "full" ;
    bf:descriptionConventions a bf:DescriptionConventions;
      rdfs:label "Resource description and access"@en.
  bf:descriptionLanguage "English".
```

Figure 3. Albert Camus BIBFRAME.

This change would allow library metadata to break free of its limitation to library discovery environments and allow it to become integrated with a wide world of information on the Web.

And while we had these shifts in communication technology, we also had major shifts in descriptive practice, from AACR1 to AACR2 to RDA and to Official RDA.

But at heart, there has been no change in how we approach access to library resources, whether the descriptions are created according to AACR2 or Official RDA, or whether we capture the information as MARC or as BIBFRAME. We create an individual metadata surrogate for every individual resource in our collections and we integrate them into similarly created descriptions for discovery.

But is this approach scalable? Given the rise of the internet and digital resources, the potential items to be added to our collections have shattered the idea of a curated collection. There are nearly an infinite number of resources available in every language on earth. Although new techniques such as entity extraction and generative Artificial Intelligence can help us with issues of scale, perhaps there is a better path we should follow.

Libraries, by nature, are conservative institutions. Although they make the shift to new technologies, they use them to recreate models with which they are comfortable:

- They used Machine Readable Cataloging, or MARC, to replace the printed catalog card.
- They used BIBFRAME to replace the MARC record.
- They look to new tools such as entity extraction and generative Artificial Intelligence to help them with issues of scale in their traditional catalog models.

But they fail to question whether the system of creating bibliographic surrogates for every item in their collections is the real problem; whether this conceptual framework itself is what should be replaced. Is it ever possible to achieve *The Brisbane Declaration's* goal of the elimination of information poverty and exclusion by following our current practices?

3. Challenges Posed by Linked Open Data

One alternative approach would be to explore the connections between facts captured in library resources themselves through the use of linked open data and avoid the creation of a catalog record completely. As an example, I'd like to briefly talk about a project at Stanford called Know Systemic Racism, or KSR. The goal of KSR is well defined: to gather and share data that show how the interconnected systems of Racism work against the Black community in the State of California.

In September of 2020, Stanford began its first ambitious project in response to the global Black Lives Matter movement, an exhibit called *Say Their Names – No More Names*. KSR is an out-growth of this first effort. It is designed to help people discover factual data about interconnected systems that pose threats to people of African descent in the United States.

The data within KSR is derived from Police Manuals, California legislation, and news events. It would be possible for the library to catalog each of the individual resources collected for KSR and create metadata surrogates for them, but the power of the collection lies in the interconnection of the data contained within the resources, not the limited and arbitrary metadata we might assign.

The graph data structure reveals the relationships, or links, between nodes. This linking structure makes it possible to connect different sources and resources that would otherwise be held apart within the library. Stanford Libraries is collecting California legislation records and California law enforcement agency policy manuals. What effect do these laws and policies have on our society? On the lives of individuals? The knowledge graph helps reveal those connections.

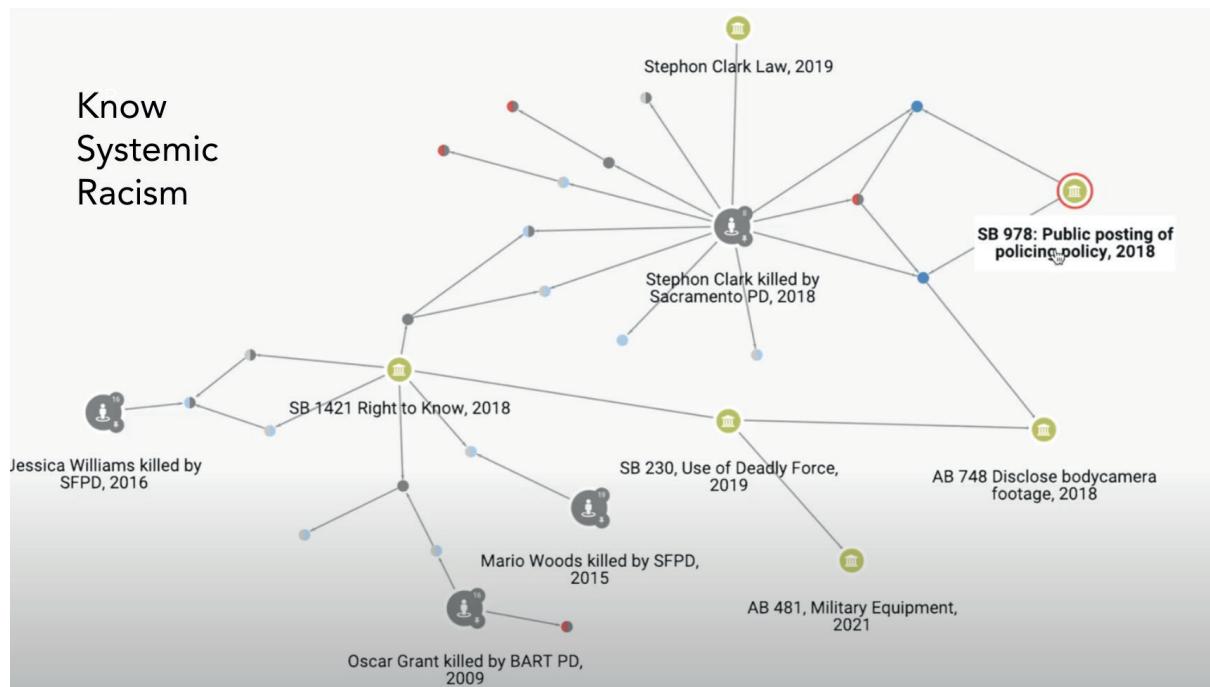


Figure 4. Know Systemic Racism as an example of Linked Open Data

Near the center of the graph, you can see the node representing the murder of Stephon Clark by the Sacramento police in 2018. By following the links in the graph, you can discover statements in police policy manuals that justified the force used in the killing, legislation created in response to the killing, such as the Stephon Clark Law in 2019 or the need to Disclose Body camera Footage in 2018, and other similar killings by Police Officers.

By relying simply on metadata created for each individual resource in this collection, you would never discover how intimately key data points in those resources relate to each other.

4. Challenges Posed by Artificial Intelligence

Another intriguing possibility is created by the combination of data disaggregation combined with the power of generative Artificial Intelligence. By aggregating data from many sources, as is traditional in library catalogs, we may obscure truths about underrepresented groups and perpetuate the myth of the majority. By disaggregating data, however, we can help researchers, community members, and policy makers better understand the needs of marginalized groups.

One experiment we have tried at Stanford is to focus a chatbot on a single repository, in this case, Black @ Stanford. The Black @ Stanford Anthology, a collaboration of Stanford's Black Community Services Center and the Stanford Archives, has the goal of gathering archival documentation and information to showcase the history of Black activism and community at Stanford. By working with an open-source version of a Chatbot, and by focusing it to look exclusively at a source of your choosing, in this case Black@Stanford, you can ensure that the voices of those in that archive are highlighted.

By accessing the search interface, anyone is able to ask questions of the interviews with Black Stanford students and faculty from the University Archives. And so, when you ask a generic question like “What was it like in 1968 at Stanford,” you get an answer derived from those interviews.

Ask the Black@Stanford Exhibit

Enter your question here

What was it like in 1968 at Stanford?

In 1971, Charles James Ogletree Jr. arrived at Stanford as a freshman and interacted primarily with other African American students in Roble Hall. The civil rights movement and the need to implement the rights granted in the Civil Rights Act of '64 and the Voting Rights Act of '65 influenced the university and students, who were absorbed with the need to push forward and solidify what had come through that legislation. The notion of diversity was also in the air, and the idea of affirmative action was very strong in the climate of that particular time.

Sources:

https://stacks.stanford.edu/file/druid:ty043xh1985/ty043xh1985_script.pdf, https://stacks.stanford.edu/file/druid:qm458gt7082/qm458gt7082_script.pdf

Figure 5. Black@Stanford as an example of Generative Artificial Intelligence

By using this technique, we can help researchers, community members, and policy makers better understand the needs of all communities by highlighting the voice of those underrepresented groups. By focusing on open data and data disaggregation, we can help mitigate information privilege and sharpen the focus on social justice and information integrity as highlighted by *The Brisbane Declaration*.

5. Conclusion

We are at a crossroads in the world of traditional metadata standards. New tools such as Artificial Intelligence and linked open data have given us opportunities to turbocharge our traditional metadata practices. Far from taking our jobs, they can make impossible jobs possible once again. But is that the direction in which we should move? The creation of metadata surrogates for every individual resource in our collections is an impossible task given the overwhelming amount of information available to us. Traditional cataloging has reached its limit.

Instead of using new tools to support old models, we should break free from those traditional methods and move into a new world of discovery based on the best opportunities those new tools offer us. Let us not make our repeated mistake of using a new technology to recreate the old world of catalog records as we move into a new world based on Artificial Intelligence and linked open

data. Faced with the power these tools bring to the World of Information Technology, what guidance can be brought from the World of Standards to make sure that they are used in a meaningful, inclusive, and sustainable way.

References

- IFLA (International Federation of Library Associations and Institutions). 2024. *The Brisbane Declaration*. <https://repository.ifla.org/server/api/core/bitstreams/c809995e-84fe-4be4-836c-01aca-3030cab/content>.
- Nichols, Tom. 2017. *The Death of Expertise: the Campaign Against Established Knowledge and Why it Matters*. New York: Oxford University Press.
- Statement of Principles Adopted by The International Conference on Cataloguing Principles*. Paris, October 1961. https://www.ifla.org/wp-content/uploads/2019/05/assets/cataloguing/IMEICC/IMEICC1/statement_principles_paris_1961.pdf.

Digitized and born-digital cultural heritage: implications for digital humanities

Andrea Capaccioni^(a)

a) University of Perugia, <https://orcid.org/0000-0003-0018-2556>

Contact: Andrea Capaccioni, andrea.capaccioni@unipg.it

Received: 26 April 2025; Accepted: 21 July 2025; First Published: 15 September 2025

ABSTRACT

This paper critically reflects on the role and implications of the distinction between digitized and born-digital cultural objects within the field of digital humanities (DH). Although this distinction is well recognised, it does not appear to occupy a central position in current debates concerning the discipline's epistemological foundations. Through a selective critical review of key contributions in the literature, two main orientations emerge: on the one hand, an approach rooted in the tradition of humanities computing, primarily focused on data modelling methodologies; on the other, a perspective more attuned to born-digital objects, understood as new cultural practices and forms native to the digital environment. The paper emphasizes the importance of further investigating this distinction not only from a terminological perspective, but above all from a methodological and epistemological standpoint. It highlights the difference between secondary documents (digitized) and primary documents (born-digital), with the aim of contributing to a more nuanced understanding of the evolution of Digital Humanities.

KEYWORDS

Digitized cultural heritage; Born-digital cultural heritage; Digital humanities; Digital heritage.

Beni culturali digitalizzati e nativi digitali: implicazioni per l'umanistica digitale

ABSTRACT

Il contributo propone una riflessione sul ruolo e sulle implicazioni della distinzione tra beni culturali digitalizzati e nativi digitali nell'ambito delle Digital Humanities (DH). Sebbene tale distinzione sia nota, essa non sembra occupare una posizione centrale nel dibattito sui fondamenti della disciplina. Attraverso una rassegna critica di alcune posizioni significative nella letteratura, emergono due principali orientamenti: da un lato, un approccio erede dell'informatica umanistica, focalizzato sulle metodologie di modellizzazione dei dati; dall'altro, una prospettiva più attenta agli oggetti nativi digitali, intesi come nuove pratiche e forme culturali proprie dell'ambiente digitale. Il contributo sottolinea l'importanza di approfondire tale distinzione non solo sul piano terminologico, ma soprattutto da una prospettiva metodologica ed epistemologica, evidenziando la differenza tra documenti secondari (digitalizzati) e primari (nativi digitali), con l'obiettivo di contribuire a una comprensione più articolata dell'evoluzione delle DH.

PAROLE CHIAVE

Beni culturali digitalizzati; Beni culturali digitali; Umanistica digitale; Patrimonio culturale digitale.

1. Introduzione

Tra le questioni aperte che si incontrano quando si affronta lo studio del patrimonio culturale digitale (*digital heritage*) c'è la distinzione tra beni culturali digitalizzati e beni culturali nativi digitali. Una questione che riguarda gli studiosi che si occupano delle *digital humanities* (DH) ma anche chi opera negli istituti culturali (biblioteche, archivi, musei). Nel Piano nazionale di digitalizzazione (PND) italiano troviamo una conferma di questi interessi comuni. Da un lato il documento segnala molti progetti attivi legati ai beni digitalizzati e ai beni nativi digitali, sebbene la distinzione tra queste due tipologie non sempre sia così netta. Alla prima appartengono iniziative di digitalizzazione di archivi, biblioteche e collezioni museali, come nel caso del fondo fotografico Alinari o delle collezioni del Museo Nazionale Romano. Alla seconda tipologia sono riconducibili progetti di digital public history, di arte digitale e di creazione di contenuti culturali su piattaforme collaborative, nati e sviluppati direttamente in ambiente digitale. È inoltre significativo che il Piano, per definire la diversa natura dei beni digitali (nativi digitali o digitalizzati), faccia riferimento alle Digital Humanities, avvalendosi dei contributi di studiosi del settore.

Quanto esposto fino qui potrebbe far pensare che siano state individuate con adeguata chiarezza le differenze tra i beni culturali digitalizzati e quelli nativi digitali. Nell'ambito delle DH la distinzione tra le due tipologie è naturalmente ben nota. Tuttavia, da una nostra cognizione, per quanto parziale, della letteratura più rilevante in questo ambito, emerge l'impressione che la distinzione tra le due tipologie di beni digitali non sia generalmente considerata un aspetto centrale nel dibattito sui fondamenti della disciplina. Non emerge, inoltre, una netta separazione tra gli studiosi che si occupano prevalentemente di beni digitalizzati e quelli che si concentrano sui beni nativi digitali. Sebbene in alcuni casi emerga una preferenza per l'uno o l'altro ambito, la letteratura non evidenzia una divisione marcata tra i due approcci. Una distinzione più netta sembra tuttavia caratterizzare la fase iniziale della disciplina, riconducibile all'informatica umanistica, in cui l'attenzione era rivolta soprattutto ai beni digitalizzati, oppure, successivamente, con l'affermarsi delle DH nei primi anni del XXI secolo, a un interesse crescente verso i beni nativi digitali.

È utile ricordare, a questo punto, che questi diversi approcci ai beni digitali sono influenzati anche da uno stato di più generale indeterminatezza che, fin dalle origini, affligge le DH. Un campo di studi che è noto per essere alla continua ricerca di identità e attraversato da molte interpretazioni. Sull'oggetto e sulle finalità delle DH il dibattito è ancora aperto. La mancanza di una chiara identificazione dei suoi confini disciplinari rende difficile tracciarne un profilo netto. Le origini sono caratterizzate da un rapporto stretto con l'informatica. Si diffonde inizialmente l'uso dell'espressione *humanities computing*, che in Italia prende il nome di informatica umanistica (Buzzetti 2019, 72). Dai primi anni di questo secolo, soprattutto nei paesi anglosassoni, comincia ad affermarsi una nuova etichetta: *digital humanities*. Non si tratta semplicemente di un cambio di denominazione, ma di una trasformazione sostanziale che testimonia la volontà di confrontarsi in modo più ampio con le innovazioni tecnologiche emerse in quegli anni. A complicare ulteriormente il quadro contribuiscono le posizioni di alcuni studiosi, che hanno messo in discussione le etichette più diffuse e avanzato proposte alternative. Tra questi ricordiamo Lou Burnard che fa presente di essersi "divertito" a sostenere "contro Willard McCarty (e altri)", autore di un'opera rilevante come *Humanities computing* (2005), che l'informatica umanistica "non esiste" (Burnard

2001). Franco Moretti, un punto di riferimento in questo ambito di studi, si è invece dichiarato convinto che l'espressione DH non significa niente ("the term 'digital humanities' means nothing") e ha proposto per la sua specializzazione un nuovo termine: "computational criticism", anche se si è detto consapevole che tutti, compreso egli stesso, avrebbero continuato a usare per comodità l'etichetta più diffusa (Dinsman 2016). Anche Jeffrey Schnapp, docente ad Harvard e figura di spicco nel rinnovamento della disciplina, è stato anche tra i firmatari del *Digital Humanities Manifesto 2.0*, ha espresso una posizione critica nei confronti di un'etichetta definita "imperfetta" in quanto, pur contribuendo a rendere la disciplina più riconoscibile, non ne coglie appieno le finalità (Schnapp 2018, cap. *Oltre le etichette*). Secondo lo studioso, la tecnologia è parte integrante della cultura e il digitale non è solo un insieme di strumenti, bensì un elemento che si intreccia profondamente con le pratiche culturali e sociali, trasformandole e venendone trasformato. Schnapp preferisce ragionare in termini di *knowledge design*, di "progettazione del sapere", in quanto richiama una disciplina che è in grado di mettere in relazione la dimensione digitale con quella analogica, in una prospettiva che valorizza il saper fare (Schnapp 2018, cap. *Oltre le etichette*). Una posizione che non appare isolata e che, per esempio, è sostenuta da Ray Siemens che interpreta le DH come una "comunità di pratica" (*community of practice*) all'interno della quale operano coloro che si occupano di progetti incentrati sul rapporto tra discipline umanistiche e tecnologie digitali (Ciotti 2024, 26-7).

Tornando alla questione di come le DH affrontano la distinzione tra le due tipologie di beni digitali, possiamo aggiungere che più che di una contrapposizione, si potrebbe parlare di coesistenza di diverse sensibilità, ed è proprio su questo aspetto che si concentra il presente contributo. Si possono individuare infatti almeno due raggruppamenti più generali che, più che distinguersi per una preferenza tra beni digitalizzati e beni nativi digitali, si differenziano per l'approccio adottato. Alcuni studiosi pongono l'accento sulla ricerca teorica e si concentrano in particolare sulla rappresentazione digitale del bene culturale, mentre altri spostano l'attenzione verso gli oggetti stessi e le pratiche sociali e culturali che da essi scaturiscono. Si tratta di una lettura che non pretende di esaurire la complessità delle DH, ma che si propone come una ipotesi interpretativa. Pur essendo principalmente orientata a offrire spunti di riflessione sulla natura e sull'evoluzione della disciplina, la presente proposta può fornire indicazioni utili anche per guidare in modo più consapevole le scelte metodologiche, strumentali e collaborative di chi opera nel settore dei beni culturali.

2. Definire il patrimonio digitale

Prima di approfondire quanto sopra esposto, soffermiamoci brevemente su una questione terminologica: cosa intendiamo per beni culturali digitalizzati e beni nativi digitali? Possiamo definire i primi come oggetti fisici (libri, dipinti, reperti, documenti, ecc.) convertiti in formato digitale tramite processi di digitalizzazione che utilizzano scanner, fotocamere o altre tecnologie. I beni culturali nativi digitali sono invece oggetti culturali che esistono solamente in forma digitale (oggetti di arte digitali, e-book, database digitali, videogiochi, software, siti web, ecc.) e la loro creazione e fruizione avviene grazie all'utilizzo di software e di strumenti informatici.

Da alcuni anni, l'UNESCO ha chiarito che il patrimonio culturale digitale è costituito da entrambe le tipologie di oggetti:

The digital heritage consists of unique resources of human knowledge and expression. It embraces cultural, educational, scientific and administrative resources, as well as technical, legal, medical and other kinds of information created digitally, or converted into digital form from existing analogue resources. Where resources are “born digital”, there is no other format but the digital object (UNESCO 2019, 1).

Nel 2010 Ricky Erway, con l'intento di spiegare per il mondo delle biblioteche il termine “nativo digitale”, ha proposto la seguente definizione: “Born-digital resources are items created and managed in digital form” (Erway 2010, 1).¹ Nell'ambito delle DH italiane, Nicola Barbuti, con riferimento a “fenomeni Born Digital e Digitized”, ha definito il *born digital heritage*:

entità native digitali che registrano nei contenuti processi, metodi e tecniche digitali rappresentativi della co-creazione digitale delle comunità contemporanee, da salvaguardare, riusare e preservare nel tempo quale potenziale memoria storica e fonte di conoscenza per le generazioni future (Barbuti 2019, 124).

In tempi recenti, ho avanzato la proposta di una definizione di bene culturale digitale, ispirata al Codice italiano dei beni culturali e del paesaggio: “un'entità concepita e realizzata in formato digitale che presenta un interesse artistico, storico, archeologico, etnoantropologico, archivistico e bibliografico o che è testimonianza della vita materiale, sociale, spirituale, di un popolo” (Capaccioni 2022, 34).

Particolarmente interessante è la presentazione di “patrimonio culturale digitale” contenuta nel PND. Il Piano ricorda che, già dal 2014, il Consiglio dell’Unione Europea ha riconosciuto il patrimonio culturale digitale come una categoria autonoma, comprendente sia i beni nativi digitali sia quelli digitalizzati. Questo riconoscimento permette di superare “la funzione ancillare del bene digitale come replica o copia dell’originale fisico e afferma la legittimità di un percorso di conoscenza autonomo, peculiare e connotato da originalità. Originalità che non discende dall’oggetto, ma dalla relazione intellettuale da cui il bene digitale prende forma e da cui attinge nuovi significati trasmissibili e non solo ‘pensabili’ ” (Istituto Centrale per la Digitalizzazione del Patrimonio Culturale – Digital Library 2023, 15-16).

3. Approcci diversi

Nelle pagine precedenti è stato osservato che, nel riflettere sulla distinzione tra beni culturali digitalizzati e nativi digitali, nelle DH sono emersi due approcci all'interno dei quali non prevale una predilezione per i beni digitalizzati o per quelli nativi digitali, ma un differente modo di intendere le finalità della disciplina. Un approccio è più concentrato sulla metodologia di cui si avvale la disciplina per esaminare e usare gli oggetti digitali: la rappresentazione degli oggetti culturali attraverso le tecnologie digitali (*modelling*). Questo “fondamento metodologico” consiste, secondo Fabio Ciotti, nella “condivisione di un insieme di metodi computazionali e linguaggi digitali adottati per la ‘modellizzazione’ dei fenomeni e dei processi appartenenti tradizionalmente al

¹ Il termine *digital object* è attestato anche in altri ambiti disciplinari, mi limito a segnalare la riflessione filosofica di (Hui 2012).

dominio delle scienze umane” (Ciotti 2024, 24). Nell’altro riscontriamo una maggiore attenzione verso l’oggetto di studio della disciplina, ovvero i prodotti dell’attività culturale, di cui si vuole valorizzare la varietà.

Proponiamo ora una rassegna, necessariamente parziale, di alcuni studiosi delle DH, con l’obiettivo di offrire degli esempi.

Il primo approccio, come abbiamo visto, è più teorico. L’attenzione al processo computazionale e alla sua capacità di trasformare il rapporto con il bene culturale è nato e si è sviluppato in quella fase in cui l’*informatica umanistica* occupava il centro della scena del dibattito sul rapporto tra scienze umane e tecnologia. La rappresentazione computazionale diventa l’interesse principale, rendendo così la natura del bene culturale (digitalizzato e nativo digitale) non fondamentale, anche se si può cogliere una preferenza per il primo tipo. Un esempio significativo di questo approccio si trova nell’opera collettiva *The Humanities computing yearbook 1989–90*, una raccolta di saggi pubblicata nel 1991 che documenta lo stato dell’arte della disciplina. Il volume, composto da oltre settecento pagine, è in gran parte dedicato a progetti legati a beni digitalizzati: corpora linguistici, progetti di *text encoding*, ricerche in campo storico e archeologico, ecc. Un breve paragrafo è riservato alla scrittura creativa (*creative writing*), a esperimenti di *fiction* e *poetry generation* e ai giochi elettronici (*The Humanities computing yearbook 1991*, 67-71).

In questa direzione, con un approccio più teorico, possiamo collocare il lavoro di Tito Orlando, tra i pionieri dell’informatica umanistica italiana, che in un suo noto saggio dedicato all’introduzione della disciplina pubblicato nel 1990 esponeva il suo punto di vista, in seguito più volte ribadito:

Potremmo dividere con chiarezza le applicazioni [delle macchine informatiche nell’ambito delle discipline umanistiche] in due settori fondamentali: il trattamento dei testi e il trattamento dei dati. Per testi intendiamo un oggetto di trattamento che sia già in qualche modo una espressione (comprendendo sia quella letteraria, che quella visiva, musicale ecc.) e che venga codificato per intero mediante un codice che rispetti per quanto possibile tutte le valenze di tale espressione. Per dati (qui l’espressione è intesa in senso convenzionale, diverso da quello tipico dell’informatica; ma siamo ora in una specie di campo neutro) intendiamo gli oggetti di trattamento che sono desunti direttamente dalla realtà per mezzo di procedimenti di individuazione e formalizzazione che precedono e determinano il procedimento di codifica per l’automazione. Il trattamento di testi è tipico delle discipline linguistiche e di critica e storia letteraria, artistica e anche musicale. Esso fornisce automaticamente (o in certi casi sono in corso studi affinché fornisca automaticamente) indici di parole, lessici, concordanze, thesauri, edizioni critiche, analisi di immagini, analisi di suoni ecc. Il trattamento di dati è tipico delle discipline storiche, e fornisce automaticamente (in maniera più o meno flessibile e sofisticata) informazioni sulle relazioni logiche fra i dati o sintesi e statistiche basate sulle loro caratteristiche (Orlando 1990, 109).

Nell’analizzare l’oggetto di studio della disciplina, attraverso un confronto esclusivo con l’informatica (non sarebbe potuto essere diversamente considerando gli anni in cui il brano è stato scritto), Orlando presenta in prevalenza esempi di testi e dati analogici (lessici, concordanze, documenti, fonti, ecc.) che vengono sottoposti a un “procedimento di codifica”, sono cioè convertiti in un linguaggio (codice) che li renda leggibili dai computer (le “macchine informatiche”). L’approccio esposto nel brano risulta parte integrante di un impianto teorico più ampio, elaborato dallo studioso per interpretare le interazioni tra discipline umanistiche e informatica.

La monografia dell'inglese Willard McCarty *Humanities computing* (McCarty 2005), già citata in precedenza, è considerata uno dei momenti più maturi della riflessione teorica dell'informatica umanistica. Nell'opera è riservato ampio spazio all'oggetto di studio della disciplina e si fa riferimento principalmente alla teoria del *modelling*. Tuttavia lo studioso sembra porre l'accento sull'uso del computer e di come sia in grado di influenzare il nostro rapporto con i beni culturali. In tempi più recenti, in ideale continuità con Orlandi, possiamo indicare il gruppo di studiosi che hanno contribuito (tra questi c'è lo stesso docente romano) alla raccolta di saggi *Digital humanities. Metodi, strumenti, saperi* (2023). Fabio Ciotti, il curatore, scrive nell'introduzione che lo scopo dell'opera è “rappresentare lo stato dell'arte di un campo di studi relativamente nuovo e in forte espansione” e dare spazio “ad autori e prospettive molteplici e diverse” (Ciotti 2023, 19). Nel volume non troviamo contributi specifici dedicati all'approfondimento della natura degli oggetti di studio. Tuttavia, nonostante una consistente attenzione riservata ai beni digitalizzati, in alcuni saggi sono presenti riflessioni dedicate a beni o prodotti culturali nativi digitali come banche dati, risorse elettroniche, biblioteche digitali, pubblicazioni open access, digital art, digital public history, ecc.

Anche nel volume di Pierre Mounier *Les humanités numériques*, tra le opere che in Francia e nei paesi francofoni hanno più contribuito alla diffusione delle DH, non abbiamo rintracciato riferimenti esplicativi al rapporto tra beni digitalizzati e nativi digitali (Mounier 2018). Si osserva, tuttavia, che gran parte del lavoro dello studioso francese si concentra sulla digitalizzazione di testi, immagini e oggetti preesistenti, mentre i riferimenti a oggetti nativi digitali risultano meno frequenti. A proposito di questi ultimi, Mounier si sofferma sul lavoro dell'artista e studiosa americana Johanna Drucker e sulle sue riflessioni dedicate alla creazione e all'analisi di oggetti culturali nativi digitali.

Il riferimento alla Drucker ci offre l'occasione per introdurre alcuni esempi relativi all'altro approccio, che si distingue per una maggiore attenzione all'ambiente digitale e nello specifico alla realizzazione di progetti basati sulle tecnologie digitali. Tutto ciò ci ricorda come il mondo accademico anglosassone abbia svolto, e continui a svolgere, un ruolo centrale nell'evoluzione delle DH. In questo ambito, il riferimento agli oggetti nativi digitali è più frequente. Nel volume intitolato *Digital_Humanities* (2012), uno dei contributi rappresentativi della nuova fase, curato da studiosi statunitensi (e tra questi Jeffrey Schnapp e Johanna Drucker), viene esposta una visione che possiamo definire più articolata della disciplina:

L'Umanistica Digitale propone nuovi modelli e strutture istituzionali per fini della ricerca, insegnamento e condivisione delle risorse intellettuali in forma collaborativa e trans-disciplinare, con l'ausilio di tecniche computazionali. L'Umanistica Digitale offre strumenti e pratiche per produrre cultura in un mondo post-stampa, ovvero in un contesto nel quale la conoscenza non si limita alla parola scritta. Preso atto che l'introduzione del digitale ha causato trasformazioni profonde a livello tecnico e culturale, va chiarito che l'Umanistica Digitale non si limita a studiare la neo-cultura digitale e nemmeno ad aggiornare il sapere tradizionale, digitalizzando documenti e testi analogici. Detto altrimenti, il digitale non rappresenta il “cosa” e il “come” dell'Umanistica Digitale. Esso ha ridefinito molte pratiche consolidate. Tra le varie discipline accademiche, l'Umanistica Digitale è stata tra le prime ad accettare la sfida, riconoscendo le possibilità e le limitazioni del digitale (Burdick et al. 2012, 195).

Gli autori identificano in modo chiaro i due principali oggetti di studio delle DH, anche se non li limitano solo a questi: la “neo-cultura digitale”, che produce oggetti nativi digitali, e il “sapere tradizionale”, che si avvale invece della digitalizzazione per studiare i beni culturali analogici. La disciplina non si limita a digitalizzare contenuti analogici o a studiare la cultura digitale, ma ridefinisce pratiche consolidate, offrendo strumenti per produrre cultura in un contesto profondamente trasformato. Le DH sono chiamate a sviluppare nuovi modelli di ricerca, insegnamento e condivisione della conoscenza, basati sulla collaborazione e sull’interdisciplinarità, grazie all’uso di tecnologie computazionali.

In tempi più recenti, Johanna Drucker è tornata sulla questione e alla domanda “What is digital humanities?” ha risposto in modo più esplicito che la disciplina si occupa di “materiali umanistici” (*humanities materials*) sia analogici che nativi digitali:

Digital humanities work is done at the intersection of computational methods and humanities materials. The research materials may be analog or born digital, and the integration with computational methods depends upon decisions at every stage of a project’s design (Drucker 2021, [1]).

Concludiamo questa rassegna, con alcune osservazioni di Lev Manovich, studioso e artista digitale, che mi pare siano in grado di cogliere bene come si sia sviluppata nel corso degli anni l’attenzione verso i beni nativi digitali:

In the humanities, the term *digital humanities*, first introduced in 2003, was becoming known. But here as well, computational analysis of cultural datasets was only one part of this emerging larger field, which also included digitization and publication of historical collections, using digital tools in teaching, and other activities. The term “digital humanities” also had another limitation for me: digital humanities scholars were working almost exclusively with datasets of historical literary texts as opposed to other types of media or contemporary digital culture—so this term was both too broad and too narrow. Finally, I already knew of a number of very impressive artistic and design projects that visualized large cultural datasets, but this work was done outside the academic departments and also did not have a single term describing it (Manovich 2020, 7).

Nel presentare le caratteristiche di un ambito di studi denominato *cultural analytics*, Manovich ci ricorda, tra le altre cose, che in origine le DH si sono prevalentemente occupate di *dataset* di testi letterari, ovvero di raccolte organizzate di testi letterari digitalizzati ad esempio di classici o di documenti antichi, piuttosto che di altri tipi di media o della cultura digitale contemporanea.

4. Una questione aperta

La distinzione tra beni culturali digitalizzati e nativi digitali è dunque una questione aperta. Per affrontarla in modo efficace, è necessario approfondire la comprensione degli aspetti costitutivi dell’oggetto digitale e analizzare in maniera più sistematica le differenze tra le due tipologie, adottando metodologie capaci di coglierne la specificità. La definizione stessa di oggetto digitale apre uno spazio di riflessione che può essere arricchito da contributi provenienti da ambiti disciplinari

diversi, come la biblioteconomia, l'archivistica, l'informatica e la filosofia. A proposito della natura dell'oggetto digitale, si possono riprendere alcune riflessioni come quelle proposte da Kallinikos, Altonen e Marton (2010), che attribuisce ad esso quattro caratteristiche fondamentali: la modificabilità, l'interattività, l'apertura e riprogrammabilità, e la distribuzione. Gli oggetti digitali si distinguono da quelli analogici grazie alla loro elevata flessibilità e alla possibilità di essere modificati continuamente. Una modificabilità che si manifesta in molteplici forme, come la riorganizzazione degli elementi, l'aggiunta o l'eliminazione di componenti. Spesso, gli oggetti digitali sono progettati fin dall'inizio come contenitori dinamici, pensati per catturare e incorporare il cambiamento nel tempo. Un'altra caratteristica fondamentale è l'interattività: gli oggetti digitali offrono percorsi alternativi per attivare funzioni o esplorare l'organizzazione delle informazioni, adattandosi alle scelte dell'utente in modo contingente. Questo apre possibilità nuove e più articolate rispetto alle risposte statiche degli oggetti fisici. Gli oggetti digitali si distinguono anche per la loro apertura e riprogrammabilità. Essi sono infatti accessibili e modificabili da altri oggetti digitali, per esempio da software per l'editing o dagli esperti che possono intervenire direttamente sul codice sorgente. Questa caratteristica favorisce una profonda interazione e trasformazione degli elementi costitutivi, favorendo un'evoluzione continua e collaborativa. Da ultimo, gli oggetti digitali sono assemblati temporaneamente da funzioni e componenti distribuiti su infrastrutture e reti digitali, senza confini fisici definiti. La loro natura fluida sposta l'attenzione dai singoli elementi ai collegamenti e alle procedure di assemblaggio, rendendo la loro esistenza e funzionamento dipendenti da un contesto dinamico e distribuito.

Ad apporti teorici di questo tipo, dovranno seguire approfondimenti sui diversi tipi di oggetti digitali, con particolare attenzione a quelli *born digital*. Potrà pertanto essere utile ripartire dalla differenza tra documenti secondari e primari. Una scansione digitale di un manoscritto medievale conservato in una biblioteca rappresenta un documento secondario, in quanto costituisce la copia digitale di un oggetto fisico preesistente. Al contrario, un'opera di net art (o Internet art) interattiva concepita per essere fruibile esclusivamente online, oppure un archivio di post pubblici su una piattaforma social, sono esempi di documenti primari: nascono direttamente in ambiente digitale e non esistono in forma analogica.

Ulteriori indagini sui beni culturali digitalizzati potranno concentrarsi, per esempio, sulle tecniche di conversione dal formato analogico a quello digitale, sulle potenzialità offerte dalla copia digitale per l'analisi e lo studio degli originali, nonché sulle questioni legate all'accessibilità e alla fruizione dei contenuti digitali. Anche le ricerche sui beni culturali nativi digitali possono svilupparsi lungo traiettorie diverse, grazie alla loro natura dinamica, interattiva e in continua trasformazione. Un ambito particolarmente promettente è rappresentato dallo studio delle pratiche culturali digitali e delle comunità che le generano e le rielaborano. A questo si affianca lo sviluppo di strumenti computazionali capaci di analizzare oggetti digitali interattivi, come le opere di net art o gli ambienti virtuali, valorizzandone le specificità formali e funzionali.

Le questioni finora affrontate, rappresentative solo di una selezione tra i molti temi possibili, richiedono ulteriori approfondimenti. In questa prospettiva, non mancano contributi significativi da parte di autorevoli studiosi, che offrono spunti per lo sviluppo futuro della ricerca. Tra questi, assumono particolare rilievo le riflessioni del filologo e bibliografo statunitense George Thomas Tanselle sul "futuro dei documenti primari":

Occuparci del futuro dei documenti primari non ci lega a nessuna forma particolare ma a qualunque forma i documenti prendano: le forme prese in passato e le forme che prenderanno in futuro. Hanno sempre preso forme varie e probabilmente prenderanno sempre una molteplicità di forme, anche se le dominanti cambiano col tempo. Le forme elettroniche sono ora all'inizio della loro supremazia, proprio come (almeno per certi scopi) le forme tipografiche cinque secoli fa (Tanselle 2004, 121-2).

Il ragionamento di Tanselle propone un principio fondamentale: l'attenzione verso i documenti primari non deve essere vincolata a una specifica forma materiale, ma abbracciare le forme che essi hanno assunto nel passato e che potranno assumere in futuro. Lo studioso ci fa capire come la storia dei documenti sia caratterizzata da una costante varietà di supporti e modalità di trasmissione, con forme dominanti che cambiano nel tempo. Oggi, le forme elettroniche sono ormai emerse come nuove protagoniste, analogamente a quanto accadde con la stampa tipografica secoli fa. Tuttavia, questa supremazia non esclude la coesistenza di molteplici forme documentarie, né riduce il valore delle forme precedenti. L'essenza del discorso di Tanselle è che la cura e lo studio dei documenti primari devono adattarsi all'evoluzione dei supporti, mantenendo come obiettivo la comprensione e la conservazione della loro molteplicità e ricchezza.

Infine, a conclusione di questa breve disamina, è importante non trascurare gli elementi comuni che accomunano le due tipologie di beni culturali digitali. Tra questi, pur potendone qui solo accennare, riveste un'importanza centrale la questione della conservazione a lungo termine, un tema cruciale per garantire la durabilità e l'accessibilità futura di tali patrimoni.

In conclusione, è ragionevole affermare che la distinzione tra beni digitalizzati e nativi digitali non debba essere relegata a una mera questione terminologica, ma possa offrire spunti rilevanti per una più profonda comprensione della natura e dell'evoluzione delle Digital Humanities. Come è emerso nel corso dell'analisi, entrambe le tipologie di beni digitali sono presenti nei principali orientamenti della disciplina, seppur con accenti differenti. Tale distinzione sembra inoltre orientare verso l'adozione di metodologie specifiche: mentre i beni digitalizzati si prestano più facilmente a tecniche di rappresentazione e modellizzazione, i beni nativi digitali richiedono approcci capaci di cogliere le pratiche culturali e sociali che li generano e li trasformano.

Riferimenti bibliografici

- Barbuti, Nicola. 2019. "Ripensare i formati, ripensare i metadati: prove tecniche di conservazione digitale." *Umanistica digitale* 5: 121-38. <https://doi.org/10.6092/issn.2532-8816/9055>.
- Burdick, Anne, Johanna Drucker, Peter Lunenfeld, Todd Presner e Jeffrey Schnapp. 2012. *Digital_Humanities*. Cambridge: MIT press.
- Buzzetti, Dino. 2019. "Alle origini dell'Informatica Umanistica: Humanities Computing e/o Digital Humanities." In *Il museo virtuale dell'informatica archeologica*, a cura di Paola Moscati e Tito Orlando. *Atti della "Segnatura"* (13 dicembre 2017). *Rendiconti dell'Accademia Nazionale dei Lincei*, Classe di Scienze morali, storiche e filologiche, S. 9, 30.1-2: 71-103. <http://web.dfc.unibo.it/buzzetti/dbuzzetti/pubblicazioni/originIU.pdf>.
- Burnard, Lou. 2001. "Dalle 'due culture' alla cultura digitale: la nascita del demotico digitale." *Bollettino '900* 1 (giugno). <http://www.unibo.it/boll900/numeri/2001-i/W-bol/Burnard/Burnard-testo.html>.
- Capaccioni, Andrea. 2022. *Umanistica digitale. Tra transizione tecnologica e tradizione*. Santarcangelo di Romagna: Maggioli editore.
- Ciotti, Fabio. 2023. "Introduzione." In *Digital humanities: metodi, strumenti, saperi*, a cura di Fabio Ciotti, 19-47. Roma: Carocci editore.
- Dinsman, Melissa. 2016. "The digital in the humanities: an interview with Franco Moretti." *LARB. Los Angeles Review of Books*. <https://lareviewofbooks.org/article/the-digital-in-the-humanities-an-interview-with-franco-moretti/>.
- Drucker, Johanna. 2021. *The Digital humanities coursebook. An introduction to digital methods for research and scholarship*. Abingdon, Oxon; New York: Routledge.
- Erway, Ricky. 2010. "Defining born digital. An essay." <https://www.oclc.org/content/dam/research/activities/hiddencollections/borndigital.pdf>.
- Hui, Yuk. 2012. "What is a digital object?" *Metaphilosophy* 43 (4): 380-95. <https://www.jstor.org/stable/24441843>.
- Istituto Centrale per la Digitalizzazione del Patrimonio Culturale – Digital Library. (2023). *Piano nazionale di digitalizzazione del patrimonio culturale (Versione 1.1)*. Ministero della Cultura. <https://digitallibrary.cultura.gov.it/il-piano/>.
- Kallinikos, Jannis, Aleksi Aaltonen, and Attila Marton. 2010. "A Theory of Digital Objects." *First Monday* 15 (6). <http://www.uic.edu/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/3033/2564>.
- Lancashire, Ian, ed. 1991. *The Humanities computing yearbook 1989-90*. Oxford: Clarendon Press.
- Manovich, Lev. 2020. *Cultural Analytics*. Cambridge: MIT Press.
- McCarty, Willard. 2005. *Humanities Computing*. London: Palgrave Macmillan.
- Mounier, Pierre. 2018. *Les humanités numériques: Une histoire critique*. Paris: Éditions de la Maison des sciences de l'homme. <https://books.openedition.org/editionsmsh/12006>.

Schnapp, Jeffrey. 2018. *Digital humanities*. Milano: Egea. Kindle.

Tanselle, George Thomas. 2004. “Il futuro dei documenti primari.” In *Letteratura e manufatti*, 121-2. Firenze: Le Lettere.

UNESCO. 2019. “Charter on the preservation of the digital heritage.” <https://unesdoc.unesco.org/ark:/48223/pf0000179529>.

Digital Identity Management and library services access: preparing for the future

Valdo Pasqui^(a)

a) Formerly University of Florence, <https://orcid.org/0000-0002-4105-2208>

Contact: Valdo Pasqui, valdopasqui@tiscali.it

Received: 31 March 2025; Accepted: 02 July 2025; First Published: 15 September 2025

ABSTRACT

Digital identities and their management systems are the core for the next generation ecosystem of trusted digital services. The European Parliament and the European Union Council have set themselves this objective with Regulation 910/2014 (eIDAS), updated in 2024, already transposed by Italy into the Code for Digital Administration (CAD). The architecture and reference framework specification started in 2021 and the ongoing implementations will lead to the adoption by 2026 of the European Digital Identity Wallet (EDIW), whose Italian instance (IT-Wallet) has been released in embryo as recent versions of the App-IO. We are therefore on the eve of a significant transformation from the current way of managing digital identities and access to digital resources and services that will impact on libraries services too. After describing the concepts and methods of access control management and digital identity with some references to current libraries context, the paper describes the new decentralized model and examines some aspects of the transformation process that all stakeholders (decision-makers, operators and users) are called upon to be aware in order to assess the difficulties and seize the opportunities offered.

KEYWORDS

eIDAS2; Digital Identity; Identity and access management; Library services.

La gestione dell'Identità Digitale e l'accesso ai servizi bibliotecari: prepararsi al futuro

ABSTRACT

Le identità digitali e i sistemi per la loro gestione sono il nucleo centrale per lo sviluppo di un ecosistema di servizi digitali affidabili ("trusted"). Il Parlamento Europeo e il Consiglio dell'Unione Europea si è posto questo obiettivo con il Regolamento 910/2014 (eIDAS), aggiornato nel 2024, già recepito dall'Italia nel Codice per l'Amministrazione Digitale (CAD). Fin dal 2021 è stata avviata la specifica dell'architettura e del framework di riferimento che con le implementazioni in corso porteranno entro il 2026 all'adozione del European Digital Identity Wallet (EDIW), Portafoglio Digitale dell'Identità, la cui istanza italiana (IT-Wallet) è stata rilasciata in embrione con le recenti versioni della App-IO. Siamo dunque alla vigilia di una significativa trasformazione rispetto alle attuali modalità di gestione delle identità digitali e dell'accesso alle risorse e ai servizi digitali che impatterà anche sui servizi bibliotecari. L'articolo, dopo aver descritto i concetti e le modalità di gestione del controllo dell'accesso e dell'identità digitale con alcuni riferimenti all'attuale contesto delle biblioteche, descrive il nuovo modello decentralizzato e esamina alcuni aspetti del processo di trasformazione sottolineando che tutti i soggetti coinvolti (decisori, operatori e utenti) sono chiamati a sviluppare la giusta consapevolezza per valutare le difficoltà e per cogliere le opportunità che si presentano.

PAROLE CHIAVE

eIDAS2; Identità Digitale; Controllo dell'accesso; Servizi bibliotecari.



1. Premessa

L'identità di una persona comprende tutto ciò che caratterizza un individuo e presenta molteplici sfaccettature se esaminata da diversi punti di vista come la filosofia, la sociologia, la medicina, la psicologia e il diritto. In psicanalisi si fa riferimento all'identità psicologica intesa come "il senso e la consapevolezza di sé come entità distinta dalle altre".¹ Nella vita di tutti i giorni ci è familiare il concetto di identità come verifica di chi sia una persona attraverso il controllo che le generalità esibite corrispondano a quelle indicate in un documento di riconoscimento valido. L'identità digitale riveste un analogo ruolo nell'ambito dei processi che governano l'accesso alle risorse e ai servizi digitali erogati dalle Pubbliche Amministrazione (PA) e dalle organizzazioni private. Tutti deteniamo una o più identità digitali, è nostra cura usarle in modo corretto, proteggerle e verificare che i servizi cui accediamo garantiscano adeguati livelli di sicurezza.

Nell'ultimo semestre del 2024 si è parlato molto dell'evoluzione della app dei servizi pubblici (App IO²) che ora consente di memorizzare la versione digitale di alcuni documenti (tessera sanitaria, patente di guida e carta europea della disabilità). Il Sistema Pubblico di Identità Digitale italiano attualmente comprende due tipologie di identità digitale, lo SPID³ e la Carta d'Identità Elettronica (CIE),⁴ con le quali i cittadini possono accedere ai servizi online delle Pubbliche Amministrazioni. App IO diventerà il portafoglio italiano dell'identità digitale (IT-Wallet), lo strumento con il quale l'Italia darà attuazione alla normativa Europea entrata in vigore nel 2024, nota con la sigla eIDAS2 (European Parliament and Council 2014). Il percorso che porterà al suo esercizio è in divenire, ma siamo alla vigilia di un profondo cambiamento rispetto alle modalità di gestione tradizionali che richiede consapevolezza e attenzione.

Anche i Musei, gli Archivi e le Biblioteche sono interessati a questa trasformazione e le biblioteche pubbliche, oltre a facilitare l'accesso ai propri servizi digitali adottando le nuove soluzioni, grazie alla loro diffusione sul territorio possono svolgere anche un importante ruolo di accompagnamento di alcune categorie di cittadini più in difficoltà a causa del *digital divide*. L'articolo richiama i termini, i principi e le modalità di gestione dell'identità digitale, descrive lo scenario attuale e le sue evoluzioni in base alla normativa europea eIDAS2 per poi esaminare alcune criticità e delineare alcune linee d'azione.

2. Gestione dell'identità digitale

2.1. Definizione e attributi

La metadattazione delle risorse bibliografiche si avvale di vari sistemi di codifica degli autori come ISNI, VIAF, ORCID, LCCN e id autore SBN che i cataloghi online propongono sotto forma di link azionabili per recuperare dati sull'autore e le sue pubblicazioni. ORCID, in coppia con una

¹ <https://www.treccani.it/vocabolario/identita/>.

² <https://ioapp.it/>.

³ Un soggetto può avere più identificatori SPID, rilasciati da fornitori diversi tra quelli attivi che sono 12: <https://www.spid.gov.it/cos-e-spid/come-scegliere-tra-gli-idp/>.

⁴ La Carta di Identità Elettronica (CIE) è emessa dal Ministero dell'Interno e prodotta dal Poligrafico e Zecca dello Stato (IPZS) <https://www.cartaidentita.interno.gov.it/la-carta/>.

password, consente anche al soggetto cui è attribuito di accedere alla propria area personale per gestire le pubblicazioni di cui è autore. La pratica corrente può indurre a ritenere che l'identità digitale sia composta da un identificatore (denominato con varie sigle: login, login-id, user, user-id, user name, nome utente, account, sign-in ecc.) accoppiato con una password. In realtà l'identità digitale è un insieme di attributi che descrivono in modo univoco un soggetto o entità in un determinato contesto.⁵ Il Codice per l'Amministrazione Digitale (CAD) (CAD 2005) all'Art.1 la definisce come “*la rappresentazione informatica della corrispondenza tra un utente e i suoi attributi identificativi, verificata attraverso l'insieme dei dati raccolti e registrati in forma digitale [...]*”.

La Tabella 1 riporta l'elenco degli attributi del Sistema Pubblico di Identità Digitale italiano nelle due varianti SPID e Carta d'Identità Elettronica (CIE).⁶ La Federazione Italiana delle Università e degli Enti di Ricerca per l'Autenticazione e l'Autorizzazione (Federazione IDEM) gestisce attributi mutuati da standard consolidati nei contesti accademici e della ricerca quali LDAPv3 (IETF 2006a) e gli schemi Cosine (IETF 2006b), inetOrgPerson (IETF 2000), eduPerson (REFEDS 2023) e Schac (REFEDS 2022).

2.2. Entità e fasi del controllo dell'accesso

Lo scenario funzionale in cui si esplicita la gestione dell'identità digitale così definita comprende alcune componenti invarianti rispetto ai modelli e alle tecnologie:

- *risorse digitali* (es. articolo di rivista, e-book, pagina web, immagini ad alta risoluzione prodotte dalla digitalizzazione di risorse analogiche, audio, video, documenti di archivio e fascicoli archivistici, dispositivi condivisi come stampanti e scanner, spazio disco in cloud, servizi applicativi ecc.);
- *entità* che richiedono l'accesso a tali risorse e che possono essere persone fisiche ma anche applicazioni, software agent, web services, interfacce programmatiche applicative (API) e dispositivi hardware;
- *sistemi preposti alla gestione dell'identità e dell'accesso* che verificano che l'entità (soggetto) richiedente sia proprio chi dichiara di essere;
- *servizi che rilasciano le identità digitali e i relativi attributi alle entità.*

La cognizione di diversi OPAC e discovery tool dei sistemi bibliotecari operanti in Italia mostra che molti di essi hanno attivato modalità di accesso controllato ai servizi, previa registrazione. Si tratta di una modalità necessaria per fornire servizi personalizzati (es. area di lavoro in cui salvare query e risultati delle ricerche, scrittura di recensioni, suggerimenti di lettura, funzionalità di tipo social per la taggatura delle risorse bibliografiche, richieste di prestito online sia locale che ILL ecc.) e per gestire l'accesso alle risorse sottoposte a vincoli (es. riviste in formato digitale con abbonamento pagato dalla biblioteca/ateneo/istituzione, visualizzazione e riproduzioni di immagini ad alta definizione, risorse digitali soggette a copyright).

⁵ Cfr. le definizioni riportate in “IT Security and Privacy — A framework for identity management” (standard ISO/IEC 24760-1:2019) e in NIST Special Publication 800-63-4: “An attribute or set of attributes that uniquely describes a subject within a given context”.

⁶ https://docs.italia.it/italia/spid/spid-cie-oidc-docs/it/versione-corrente/attributi_utente.html.

Attributo	Descrizione	Supportato da	Person Identification Data per le persone fisiche (O=Obbligatorio)
spid_code	codice identificativo univoco assegnato dal gestore dell'identità digitale	spid	personal_administrative_number
given_name	Nome	spid, CIEid	given_name (O) given_name_birth
family_name	Cognome	spid, CIEid	family_name (O) family_name_birth
place_of_birth	Luogo di nascita	spid, CIEid	birth_place (O)
birthdate	Data di nascita	spid, CIEid	birth_date (O)
			nationality (O)
gender	Sesso	spid, CIEid	sex
company_name	Ragione o denominazione sociale	spid	per le persone giuridiche (legal person): - current legal name (O) - unique identifier (dello Stato Membro) (O) - una serie di altri attributi opzionali
registered_office	Sede legale	spid	
fiscal_number	Codice fiscale della persona fisica	spid, CIEid	
company_fiscal_number	Codice fiscale Persona Giuridica.	spid	
vat_number	Partita IVA	spid	
document_details	Descrizione del documento d'identità	spid, CIEid	tre metadati obbligatori relativi su chi ha emesso i PID: - issuing_authority - issuing_country - expiry_date (dei PID)
phone_number	Numero di telefono mobile	spid, CIEid	mobile_phone_number
phone_number_verified	Valore booleano che indica se il numero di telefono mobile è stato verificato dal fornitore dell'identità digitale	CIEid	
landline_number	Numero di telefono fisso	CIEid	
email	Indirizzo di posta elettronica	spid, CIEid	email_address
email_verified		CIEid	
e_delivery_service	Domicilio digitale Indirizzo casella PEC	spid, CIEid	
eid_exp_date	Data di scadenza identità	spid	
address		spid, CIEid	- resident_address - resident_country - resident_state - resident_city - resident_postal_code - resident_street - resident_house_number
			Portrait (foto dell'utente del wallet)

Tabella 1. Confronto attributi del Sistema Pubblico di Identità Digitale italiano vs Person Identification Data in Art.3.3 di (European Commission 2024)

La logica di alto livello del controllo dell'accesso può essere descritta esaminando le quattro fasi di cui si compone:

- a) *Registrazione* – è una fase “una tantum” che consente di creare e assegnare un’identità digitale, spesso indicata come “onboarding” di un nuovo utente o “crea il tuo Account” se svolta in autonomia online;
- b) *Identificazione* – la risorsa o il servizio a cui un’entità (es. una persona) vuole accedere le chiede di farsi riconoscere;
- c) *Autenticazione* – l’identità digitale fornita viene verificata, riscontrando che l’entità che l’ha dichiarata sia proprio quella che afferma di essere;
- d) *Autorizzazione* – se l’autenticazione ha avuto esito positivo viene controllato che l’entità abbia i diritti per accedere alla risorsa o al servizio richiesti.

2.3 Assegnazione dell’identità digitale

Le modalità di registrazione e assegnazione dell’identità digitale sono un passaggio fondamentale per assicurare l’affidabilità dell’intero ecosistema di servizi basati sull’identità digitale e per minimizzare i rischi di impersonificazione (un soggetto che agisce assumendo l’identità di un altro). È bene sottolineare che le identità digitali rilasciate in modo verificato da organismi appositamente preposti che seguono procedure codificate (es. SPID e CIE) hanno un livello di affidabilità non confrontabile rispetto a quelle (auto)create online dagli utenti stessi. La registrazione e l’attivazione di SPID sono svolte da uno dei gestori attivi (Identity Provider) esibendo un documento italiano in corso di validità o di persona o tramite CIE (online o attraverso app), ma anche da remoto (con o senza operatore⁷). La CIE può essere richiesta presso un qualsiasi Comune e il rilascio comprende la verifica dei dati anagrafici e l’acquisizione delle impronte digitali e, in caso di primo rilascio, l’esibizione di un documento d’identità in corso di validità (altrimenti occorrono due testimoni). Gran parte degli OPAC e dei portali delle biblioteche italiane offrono la registrazione online che comporta l’immissione dell’identificativo utente (di solito indirizzo di posta elettronica) di una password e di alcuni dati anagrafici e di contatto. Solo alcuni sistemi bibliotecari (es. Biblioteche di ROMA, la Rete bibliotecaria di Romagna e San Marino, il Sistema Bibliotecario della Provincia di Reggio Emilia, il Polo Bibliotecario Modenese, il Polo regionale SBN Sardegna, il Polo COBIRE della Regione Toscana, il Polo MiC Venezia, il Polo bibliotecario regionale del Veneto) hanno colto l’opportunità di avvalersi dell’accesso con SPID/CIE anche per consentire la registrazione recuperando dal Sistema Pubblico d’Identità gli attributi necessari, così rendendo l’operazione completamente automatica per gli utenti e assicurando alle biblioteche il recupero di dati certificati.

⁷ Occorre evitare di far circolare copie dei documenti personali di riconoscimento perché truffatori venuti in possesso di tali copie hanno attivato un secondo SPID intestato a una persona con email e telefono diversi agendo poi per suo conto in modo fraudolento in alcuni servizi (es. INPS).

2.4 Identificazione e autenticazione: password e fattori

Le successive fasi del processo di controllo dell'accesso sono l'identificazione e l'autenticazione che si ripetono in occasione di ogni accesso a un servizio o una risorsa digitale. Questi due momenti si possono rappresentare con la sequenza: "Dimmi chi sei?" – "Sono Paolino Paperino" – "Dimostramelo!". Al termine di questa interlocuzione l'accesso richiesto può essere concesso o negato. Nel primo caso gli attributi che costituiscono l'identità digitale (eventualmente con altri attributi del soggetto/entità registrato) sono utilizzati per effettuare i controlli autorizzativi che determinano il contesto di risorse e funzioni locali abilitate all'utente.

La dimostrazione della identità digitale dichiarata consiste nell'esibizione da parte dell'utente di uno o più fattori (credenziali) classificabili nelle seguenti categorie:

- *Qualcosa che si conosce*: password, PIN;
- *Qualcosa che si possiede*: chiavetta fisica (hardware token), smartcard, codice one-time (OTP) inviato per email o SMS, codice OTP generato da un dispositivo esterno (es. RDS SecurId), codice o richiesta di conferma (es. "Autorizza richiesta") generati su una APP di un dispositivo mobile (smartphone) associato all'identità digitale e a cui si accede in modo protetto con PIN o controllo biometrico;
- *Qualcosa che si è*: una proprietà biometrica come un'impronta digitale, la scansione della retina, della faccia o della voce (spingendosi fino al DNA in contesti molto sensibili).

L'utilizzo della sola password è la modalità più diffusa ma anche la più debole (un solo fattore) e esposta a molti rischi di compromissione (password facili da scoprire, meccanismi che inducono gli utenti a comunicare le proprie password a soggetti/siti fraudolenti, malware che spiano e catturano le password usate ecc.), pertanto si sta sempre più affermando l'imposizione di due o più fattori. Esistono anche soluzioni che prevedono lo scambio di codici temporanei crittografati (token) mediante coppie di chiavi crittografiche (pubblica e privata) distribuite tramite certificati digitali secondo gli standard della Public Key Infrastructure (PKI).⁸

SPID e CIE prevedono tre livelli di utilizzo, il secondo (due fattori) è lo standard, per SPID consiste nell'invio di un OTP tramite SMS o app e per la CIE prevede l'uso del OTP o la scansione di un QRCode.

3. Modelli e soluzioni un ecosistema di servizi “trusted”

Le identità digitali e i sistemi per la loro gestione sono il nucleo centrale attorno al quale costruire un ecosistema di servizi “trusted” ovvero dei quali gli utenti si fidano, grazie alla combinazione e l'attuazione di standard, tecniche, procedure, regole e comportamenti che assicurano elevati livelli di affidabilità (reliability), verificabilità (accountability) e disponibilità (availability).

⁸ La FIDO Alliance, l'associazione di industrie ICT che sviluppa standard e soluzioni per l'autenticazione, fonda la soluzione Passkeys, supportata da Google, Amazon, Apple, Microsoft e altre aziende, proprio sulla crittografia a doppia chiave per consentire all'utente di autenticarsi al server senza l'uso di password e l'invio di altri dati: <https://fidoalliance.org/passkeys/>.

3.1 Sicurezza, comportamenti e modelli

Per quanto attiene la sicurezza nel contesto delle biblioteche è bene precisare che oltre a proteggere le risorse bibliografiche e digitali occorre anche tutelare la privacy degli utenti e le loro identità digitali. I sistemi bibliotecari non sono esenti dai rischi di attacchi con effetti rilevanti come dimostra il caso della British Library (BL 2024; Grove 2024). Anche un catalogo online o un discovery tool possono essere obiettivo di attacchi volti a es-filtrare le identità digitali e a violare la privacy degli utenti estrapolando dati sugli interessi di lettura, studio e ricerca e sugli orientamenti politici e sessuali che possono essere usati in modo diretto contro l'utente o indirettamente a fini commerciali (vendita dei dati).

Per limitare questi rischi occorre agire congiuntamente su due livelli: quello comportamentale che ricade sotto la responsabilità degli utenti e quello strutturale che concerne i modelli e le architetture dei sistemi di gestione dell'identità digitale.

Sul primo aspetto si rimanda alla lettura della sezione dedicata alla Cybersecurity del sito del Garante della Privacy che propone un ampio insieme di schede informative⁹ sulle misure di sicurezza, la generazione e la protezione delle password e le modalità per proteggersi dai tentativi di furto d'identità (phishing, smishing, vishing, ransomware) (Clusit 2025). Per mitigare i rischi di sicurezza legati all'uso delle password si sta diffondendo la pratica di utilizzare software che generano password complesse e le memorizzano in modo crittografato (Kime 2024).

La soluzione più datata, ma ancora oggi molto diffusa, per la gestione del controllo dell'accesso prevede funzioni specializzate degli applicativi e tabelle del database per registrare gli attributi degli utenti (password comprese). La maggior parte dei sistemi gestionali installati presso le biblioteche, come si riscontra in un numero molto elevato di cataloghi online e discovery tools bibliotecari, seguono questo approccio perché sono concepiti come soluzioni autonome (verticali) e dunque mettono a disposizione dei propri clienti le funzionalità per registrare gli utenti e per supportare almeno le operazioni base di cambio e recupero password dimenticata.

3.1.1 Il modello di gestione delegata

Per ovviare alla proliferazione di silos contenenti le identità digitali, fin dalla metà degli anni Novanta del secolo scorso sono stati sviluppati sistemi dedicati, denominati Identity and Access Management (IAM), che concentrano tutte le funzionalità e procedure concernenti la gestione delle identità:

- gestire il repository degli attributi delle identità digitali;
- implementare le funzioni di registrazione (self-provisioning), cambio e recupero password dimenticata;
- offrire procedure di provisioning e deprovisioning massivo utili in ambiti aziendali e istituzionali con un rilevante numero di utenti soggetti a ben definite logiche di registrazione e disattivazione (es. studenti di un ateneo, dipendenti di un ente territoriale o di un'azienda multi sede, cittadini residenti in un comune ecc.);
- implementare tecniche di autenticazione a più fattori;
- adottare le policy di sicurezza per la protezione dagli attacchi informatici;

⁹ <https://www.garanteprivacy.it/temi/cybersecurity>.

- abilitare l'autenticazione singola a più servizi, Single Sign-On (SSO);
- implementare i protocolli e gli standard previsti per abilitare la partecipazione a sistemi federati di gestione delle identità, tra cui in Italia SPID e CIE.

Questo approccio ha portato anche alla diffusione del modello di gestione delegata dell'identità digitale, Delegated Identity Management, ormai largamente adottato dalle applicazioni web e dai servizi operanti in Cloud:

- a) un sistema, denominato Identity Provider (IdP), ha in carico la gestione delle identità digitali e gestisce le operazioni necessarie per svolgere le fasi di identificazione e di autenticazione;
- b) le applicazioni e le risorse digitali sono protette da un componente, denominato Service Provider, che quando un'entità (es. utente) ne richiede l'accesso lo reindirizza attraverso un canale sicuro (crittografato) verso l'IdP;
- c) le interazioni per portare a compimento l'identificazione e l'autenticazione avvengono direttamente tra l'entità richiedente e l'IdP;
- d) se l'autenticazione fallisce l'entità richiedente viene rinviata al SP con messaggio d'errore e il SP nega l'accesso, altrimenti il SP riceve un token e degli attributi (detti asserzioni) che permettono al SP di abilitare l'accesso alla risorsa/servizio richiesto.

Le politiche e le modalità (es. autenticazione a più fattori) di controllo dell'accesso sono così centralizzate in un unico sottosistema specializzato e condiviso da più applicazioni e servizi, garantendo una gestione protetta, unificata e coordinata (regole, policy, tecniche di sicurezza ecc.) delle identità digitali degli utenti. Così i singoli software applicativi si svincolano dalle problematiche tecniche e di sicurezza demandando al sottosistema dedicato la gestione delle identità digitali.

3.1.2 Single Sign-On e la gestione federata delle identità

Il modello della gestione delegata è al centro del funzionamento del Single Sign-On e dei sistemi di Identity Management Federato. Nel primo scenario, tipicamente aziendale, quando un utente accede a un'applicazione il SP lo reindirizza verso l'IdP che verifica l'esistenza di una sessione di lavoro attiva, creata in occasione della precedente autenticazione di un'altra applicazione. In caso affermativo l'IdP restituisce al SP l'autorizzazione per concedere l'accesso senza ripetere nuovamente le fasi di identificazione e autenticazione (da qui il termine "single"). Nel secondo scenario più organizzazioni si accordano per condividere procedure formali e tecniche attraverso le quali riconoscere reciprocamente le identità digitali emesse da ciascuna di loro. Ogni aderente alla federazione gestisce il proprio IdP che è responsabile dell'assegnazione, della gestione e della verifica (autenticazione) dell'identità digitali delle proprie entità (persone, servizi ecc.). Quando un utente membro della federazione chiede l'accesso ad un servizio/risorsa condiviso/a il relativo SP lo indirizza verso l'IdP della sua organizzazione a cui sono demandate l'identificazione e autenticazione. Se l'autenticazione con il proprio IdP avviene con successo questo restituisce al SP un insieme di attributi che gli consentono di riconoscere l'utente, sbloccare l'accesso e applicare le politiche autorizzative. In tal modo le identità digitali e i fattori usati restano confinati nell'ambito dell'interazione tra l'utente e il sistema di gestione dell'identità dell'istituzione cui afferisce, riducendo il numero di identità digitali usate.

Il Sistema federato IDEM della rete accademica e della ricerca in Italia consente ai docenti, ai ricercatori e agli studenti delle Università e Istituti di ricerca italiani di accedere con la propria identità digitale istituzionale a risorse digitali condivise¹⁰ tra queste istituzioni, tra quelle aderenti alla rete mondiale EDUGain e con editori commerciali come EBSCO Publishing,¹¹ Elsevier,¹² IEEE Xplore,¹³ Casalini Biblioteca Digitale Torrossa¹⁴ e fornitori di servizi come Proquest¹⁵ e Medialibraryonline (Mlol).¹⁶

Il Sistema di Identità Digitale italiano è di fatto una federazione composta dagli attuali 12 fornitori di identità SPID e dal Ministero degli Interni che rilascia e gestisce la CIE.

L'Unione Europea utilizza l'approccio federato per garantire l'interoperabilità transfrontaliera tra gli Stati membri che adottano differenti schemi di identificazione nazionale (eID Building Block¹⁷). Ogni Paese gestisce un “nodo eIDAS” connesso punto a punto con i corrispettivi nodi di altri Stati a formare una maglia che garantisce l'interoperabilità tecnica e la mappatura tra i diversi sistemi di identificazione nazionali.¹⁸ Quando l'utente accede ad un servizio online trova il link/bottone “Login con eIDAS” (es. accesso all'Anagrafe Nazionale della Popolazione Residente – ANPR¹⁹) che gli permette di scegliere il Paese di provenienza e di essere reindirizzato al corrispondente sistema di identificazione (IdP) per poi essere automaticamente riportato al servizio richiesto²⁰ se l'autenticazione avviene con successo.

I messaggi e l'orchestrazione degli scambi tra le componenti di questi modelli si avvalgono di protocolli e linguaggi standardizzati: Security Assertion Markup Language (SAML) (OASIS Open 2008), OpenId Connect (OIDF 2023) e Oauth 2.0 (IETF 2012). Software open source come Shibboleth²¹ e Central Authentication Service (CAS)²² o proprietari come Microsoft Entra ID²³ forniscono i componenti per l'implementazione di queste architetture. OCLC supporta il modello federato attraverso il componente EZproxy²⁴ che permette l'accesso alle risorse digitali della bi-

¹⁰ Risorse della Federazione IDEM <https://www.idem.garr.it/partecipare/risorse-idem>.

¹¹ Users: Start your research => Use my location => scegliere ente/istituzione di appartenenza <https://www.ebsco.com/news-center/press-releases/ebsco-publishing-and-ebsco-information-services-merge>.

¹² Es. Elsevier ScienceDirect <https://www.sciencedirect.com/>: “SignIn” => “Find your organization”.

¹³ “Institutional Sign in” <https://ieeexplore.ieee.org/Xplore/home.jsp>.

¹⁴ <https://access.torrossa.com/it/user/access>: “Accesso Istituzionale Federato”.

¹⁵ <https://www.proquest.com>: “Sign into My Research” => “Find your institution now”.

¹⁶ <https://www.medialibrary.it/home/index.aspx>: “Accedi” => “Seleziona un Ente”

¹⁷ <https://ec.europa.eu/digital-building-blocks/sites/display/DIGITAL/eID>.

¹⁸ Mappa delle interconnessioni da e verso l'Italia: <https://eidas.ec.europa.eu/efda/browse/notification/eid-chapter-contacts/IT>.

¹⁹ <https://www.anagrafenazionale.interno.it/servizi-al-cittadino/> (Attenzione: dopo aver selezionato il paese di provenienza occorre scorrere tutta l'informativa per sbloccare il bottone e proseguire).

²⁰ Questa infrastruttura è uno dei mattoni base per la costruzione dello Sportello digitale unico (Single Digital Gateway) previsto dal PNRR. <https://www.agid.gov.it/it/agenzia/attuazione-misure-pnrr/sportello-digitale-unico-single-digitale-gateway>.

²¹ <https://www.shibboleth.net/>.

²² <http://jasig.github.io/cas/4.0.0/index.html>.

²³ Nome assegnato dal marzo 2024 a Azure Active Directory, il servizio di gestione delle identità e degli accessi in cloud (piattaforma SaaS MS Azure, MS 365, il portale di Azure) e a risorse interne (intranet aziendale) <https://learn.microsoft.com/it-it/entra/fundamentals/whatis>.

²⁴ OCLC sta promuovendo attraverso una serie di webinar questa soluzione che consente anche il monitoraggio statistico dell'accesso alle risorse: <https://www.oclc.org/en/ezproxy.html>.

blioteca integrando (mediante SAML e LDAP) i sistemi di gestione locali in modo da consentire agli utenti di usare le identità digitali delle proprie istituzioni (es. università).

3.2 La nuova frontiera: gestione dell'identità decentralizzata

Per superare i rischi di sicurezza a cui sono esposti i repository delle identità digitali sta emergendo il modello di gestione decentralizzata o “self-sovereign identity” che elimina i repository delle identità digitali le quali sono direttamente controllate dal soggetto che le detiene per mezzo di un “digital wallet” (portafoglio digitale), una app installata su un dispositivo mobile (es. smartphone) o un servizio in cloud.²⁵ In questa modalità agiscono i seguenti elementi (W3C 2025):

- *Verifiable Digital Credential (VC)* – un insieme di attributi (asserzioni) relativi a una persona, corredati da alcuni metadati che descrivono l’organizzazione che li emette (issuer), la cui provenienza è resa sicura e verificabile mediante tecniche crittografiche, hanno una data di scadenza e possono essere sospese o revocate.
- *Issuer* – organizzazione (PA, banca, università ecc.) che crea le VC, le firma con la propria chiave privata e le rilascia al proprietario/detentore (holder) dopo averlo identificato.
- *Holder* – il soggetto titolare che riceve le VC, le salva nel proprio digital wallet e le usa (presenta) per accedere a servizi e risorse digitali.
- *Digital Wallet (Portafoglio Digitale)* – lo strumento (es. app per dispositivo mobile) che permette di salvare, memorizzare e gestire le credenziali verificabili.
- *Service Provider o Relying Party* – il componente che protegge l’accesso a una risorsa o servizio digitale e che verifica le credenziali digitali presentate dall’utente per decidere se concedergli l’accesso richiesto.

Rispetto ai modelli descritti in precedenza il controllo dell’accesso non si avvale più degli Identity Provider e l’interazione avviene direttamente tra il servizio (Relying Party) e il Portafoglio digitale (Digital Wallet) che l’utente (Holder) usa per esibire le sue credenziali verificabili (VC).

L’implementazione dell’infrastruttura tecnico-funzionale si avvale della crittografia a chiave pubblica, dei protocolli definiti da OpenId per la gestione delle credenziali verificabili²⁶ e di standard per la loro codifica come JSON Web Tokens (JWTs), CBOR Web Tokens (CWTs), JSON Advanced Electronic Signature (JadES) e ISO-18013-5:2021 for Mobile Driving License(mDLs).

4. Verso un nuovo ecosistema di gestione dell'identità digitale

4.1. eIDAS2, EDIW e IT-Wallet

Dal 20 maggio 2024 è entrato in vigore il Regolamento EU 2024/1183, noto come eIDAS2 che ha aggiornato il precedente 910/2014 (eIDAS) per consolidare lo sviluppo di servizi digitali fiduciari (“trusted”) e avviare il processo di armonizzazione dell’identità digitale con l’obiettivo di ridurre i rischi di sicurezza e i costi gestionali derivanti dalle diverse soluzioni adottate dai Paesi europei

²⁵ Il concetto di digital wallet è mutuato dall’ambito del commercio e dei pagamenti elettronici.

²⁶ <https://openid.net/wg/digital-credentials-protocols/digital-credentials-protocols-dcp-working-group-repository2/>.

(European Parliament and Council 2014). Ancor prima, nel 2021, ha preso il via la definizione della Architettura e del Framework di Riferimento (European Commission 2025) dello European Digital Identity Wallet (EDIW), il portafoglio europeo elettronico dell'identità digitale. EDIW segue il modello decentralizzato ed è “un mezzo di identificazione elettronica che consente all'utente di conservare, gestire e convalidare in modo sicuro dati di identità personale e attestati elettronici di attributi al fine di fornirli alle parti facenti affidamento sulla certificazione digitale”²⁷. Un insieme di undici use case concorre a delineare lo spettro di applicazione molto ampio del portafoglio digitale: memorizzare la patente di guida in formato digitale, aprire un conto corrente bancario da remoto, salvare documenti e biglietti di viaggio, apporre firme digitali, registrare attestati e certificazioni di studio, identificarsi e autenticarsi a servizi digitali pubblici e usare identità digitali che attestano l'appartenenza a organizzazioni, enti, associazioni, aziende, catene commerciali. Per esempio le biblioteche potranno rilasciare l'equivalente digitale delle tessere analogiche di iscrizione dei propri utenti sotto forma di attestazioni elettroniche di attributi. Per ciascuno degli scenari individuati sono in corso sperimentazioni attraverso progetti pilota²⁸ e la roadmap completa delle implementazioni è in liberamente consultabile.²⁹

Il 4 dicembre 2024 la Commissione EU ha pubblicato i primi cinque Atti implementativi (Implementing regulations)³⁰ che definiscono standard, specifiche e procedure relative alla funzionalità tecniche dei wallet e al processo di certificazione necessario per garantire che i wallet siano sicuri e proteggano i dati personali degli utenti.³¹

Per quanto riguarda l'identificazione digitale il Framework di Riferimento prevede che l'accesso ai servizi pubblici e privati avvenga presentando dei Person Identification Data (PID), insiemi di dati identificativi che consentono di stabilire l'identità di una persona fisica o giuridica o di una persona fisica che rappresenta una persona giuridica. I PID sono accompagnati da metadati che identificano il fornitore responsabile dell'emissione e della revoca di questi dati, al quale spetta assicurare che il PID di un utente sia crittograficamente legato al suo wallet. Attribuiti (cfr. Tabella 1), formati e codifiche dei PID sono definiti nel Regolamento EU (European Commission 2024).

Il Regolamento eIDAS2 prescrive che ciascun Stato membro debba fornire almeno un portafoglio europeo di identità digitale entro 24 mesi dal 20 maggio 2024, il cui uso è comunque facoltativo poiché viene mantenuta la possibilità di accedere ai servizi pubblici e privati con altri mezzi di identificazione e autenticazione esistenti. La normativa inoltre stabilisce che l'emissione dei portafogli digitali, l'utilizzo e la revoca devono essere gratuite, che gli Stati Membri devono prevedere meccanismi di convalida gratuiti e che il codice sorgente dei componenti software dei wallet sia rilasciato con licenza open source.

²⁷ Cfr. punto 42) aggiunto dal Regolamento EU 2024/1183 al regolamento (UE) n. 910/2014.

²⁸ <https://ec.europa.eu/digital-building-blocks/sites/display/EUDIGITALIDENTITYWALLET/What+are+the+Large+Scale+Pilot+Projects>.

²⁹ <https://github.com/eu-digital-identity-wallet/eudi-wallet-reference-implementation-roadmap>.

³⁰ <https://ec.europa.eu/digital-building-blocks/sites/display/EUDIGITALIDENTITYWALLET/EU+Digital+Identity+Wallet+milestone+reached+as+Commission+adopts+implementing+regulations>.

³¹ Il 26 giugno 2025 sono stati rilasciati in pubblica consultazione fino al 15 luglio ulteriori 3 Atti: <https://ec.europa.eu/digital-building-blocks/sites/display/EUDIGITALIDENTITYWALLET/New+batch+of+Implementing+Acts+published+for+feedback>.

A livello normativo l'Italia ha recepito il regolamento eIDAS2 con la Legge 29 aprile 2024 che ha aggiunto al CAD il nuovo l'Art. 64-quater "Sistema di portafoglio digitale italiano – Sistema IT-Wallet". Al livello pratico la versione corrente della AppIO consente di scaricare e memorizzare la versione digitale della patente di guida, della tessera sanitaria e la Carta Europea della Disabilità³² e rappresenta la prima istanza dell'IT-Wallet. Al 28 giugno 2025 gli utenti di AppIO attivi negli ultimi 12 mesi erano 12 milioni (con 4.4 milioni di patenti di guida digitali, 4.8 milioni di tessere sanitarie e 0.1 milioni di Disability Card scaricate).³³ In tempi successivi sono pianificati il caricamento della CIE, del passaporto, della tessera elettorale, del certificato di nascita e di titoli di studio. Entro il 2026 (scadenza del PNRR) è pianificata l'integrazione con EDIW.³⁴

4.2 Prepararsi al cambiamento tra luci e ombre

Il percorso per lo sviluppo e l'avvio di EDIW e di IT-Wallet è in itinere. Le Linee Guida AgID³⁵ per "disciplinare gli standard tecnologici e le soluzioni tecniche ed organizzative" la cui pubblicazione secondo l'Art. 64-quater comma 3 del CAD doveva avvenire entro sessanta giorni dall'entrata in vigore (1 Maggio 2024) ancora non sono state pubblicate. Questo ritardo avrà un impatto sul processo di attuazione soprattutto per quanto riguarda il coinvolgimento del settore privato per la produzione di portafogli digitali (wallet) e per la fornitura/emissione di attestati elettronici di attributi. Gli operatori economici devono infatti disporre di un quadro di riferimento chiaro (regole e procedure) dal momento che queste attività, sempre in base all'Art. 64-quater, sono sottoposte a accreditamento.³⁶

Predisorsi alla trasformazione implica che tutti i soggetti coinvolti (privati cittadini, PA centrali e locali, operatori commerciali) acquisiscano la consapevolezza delle opportunità e delle complessità che questa evoluzione comporta.

Il principale vantaggio del Sistema IT-Wallet è la riduzione delle criticità di sicurezza e degli oneri di protezione degli attuali sistemi di gestione delle identità digitali basati sulle password. Per il momento è prematuro prevedere per quanto tempo, dopo il 2026, l'ecosistema di gestione delle identità digitali manterrà operative tutte e tre le modalità SPID, CIE, IT-Wallet. L'autenticazione tramite CIE confluirà nel IT-Wallet che tuttavia non è uno strumento obbligatorio. La sostenibilità economica avrà un peso significativo sull'assestamento degli scenari futuri.³⁷

L'adozione e la diffusione dell'IT-Wallet dovrà essere accompagnata da un efficace e ampia attività

³² <https://ioapp.it/documenti-su-io/>.

³³ <https://ioapp.it/numeri>.

³⁴ I soggetti attuatori sono PagoPA e IPZS, di concerto con AgID e l'Autorità Nazionale per la Cybersicurezza (ACN).

³⁵ <https://www.agid.gov.it/it/piattaforme/it-wallet>.

³⁶ Una prima versione delle Specifiche Tecniche è disponibile a: <https://italia.github.io/eid-wallet-it-docs/versione-corrente/en/index.html>. Queste forniscono raccomandazioni, requisiti e risorse relative al sistema IT-Wallet e quando saranno formalmente adottate insieme alle Linee Guida costituiranno il quadro di riferimento che regolerà il Sistema IT-Wallet.

³⁷ Con il comunicato del 5 marzo 2025 (<https://innovazione.gov.it/notizie/comunicati-stampa/identita-digitale-firma-to-il-decreto-che-assegna-40-milioni-di-euro-ai-gestori-spid>) il Sottosegretario alla Presidenza del Consiglio con delega all'Innovazione, Alessio Butti, ha annunciato la firma del Decreto che assegna 40 milioni di euro agli operatori impegnati nell'erogazione di SPID per garantirne "continuità e sostenibilità, comunque due di loro (Aruba e InfoCert) per lo SPID ad uso personale, dopo il primo anno gratuito, hanno introdotto un costo dal secondo anno.

di informazione, sensibilizzazione e promozione delle politiche di adozione. I decisori politici delle PA locali dovranno percepire questa trasformazione non come un mero adempimento da attuare al livello minimo indispensabile, ma come una scelta strategica per la costruzione dell'ecosistema di servizi digitali affidabili che è la finalità ultima di eIDAS2. IT-Wallet, evoluzione di App IO, sarà il portafoglio digitale conforme a EDIW rilasciato dallo Stato italiano come previsto dalla normativa italiana, ma gli operatori commerciali dovranno essere stimolati e agevolati per sostenere gli investimenti necessari affinché l'adozione del portafoglio non resti confinata al solo ambito dei servizi delle PA. A tal fine sarà necessario individuare e fissare criteri che consentano lo sviluppo di un modello commerciale per il rilascio dei portafogli digitali conformi a EDIW e che rispetti le prescrizioni del Regolamento eIDAS2 sulla gratuità di alcuni servizi. Questa non è un'incongruenza poiché la normativa circoscrive il contesto assegnando agli Stati il compito di fornire strumenti gratuiti per verificare la validità e l'autenticità degli EDIW, per verificare l'autenticità e la validità delle identità digitali delle "relying parties", stabilisce che per le persone fisiche ("natural person") le firme digitali e le operazioni di emissione, uso e revoca degli EDIW siano gratuite. Si tratterà dunque di disegnare uno scenario che da un lato non comporti costi per i privati cittadini (persone fisiche) e dall'altro assicuri la sostenibilità economica dell'intero sistema attraverso modalità di tariffazione da applicare all'utilizzo da parte dei professionisti e delle aziende, differenziando tra l'uso di attestazioni qualificate e non qualificate e tenendo conto della tipologia dei soggetti che le emettono e dei servizi cui si vuole accedere.

Infine, i cittadini dovranno essere accompagnati affinché il portafoglio di identità digitale non sia destinato solo agli utenti tecnicamente più evoluti. Per questo sarà necessario dedicare una particolare cura all'usabilità dei wallet, alla formazione e al supporto agli utenti prevedendo "punti di facilitazione digitale" diffusi sul territorio³⁸.

Le biblioteche possono cogliere l'opportunità di inserirsi in questo percorso evolutivo con più ruoli: avvalersi delle attestazioni digitali e dei PID per aggiornare i propri sistemi di controllo dell'accesso al fine di ridurre il rischio di attacchi informatici, rilasciare attestazioni digitali dell'iscrizione ai propri utenti fruibili in vari circuiti (es. eventi culturali, spettacoli, librerie ecc.), svolgere funzioni di reference per i cittadini. Le biblioteche che hanno già integrato i servizi digitali con SPID e CIE sono avvantaggiate nel cambiamento poiché in gran parte si sono avvalse di sistemi federati di ambito territoriale (es. FedERa per l'Emilia Romagna, le analoghe infrastrutture della Regione Toscana e della Regione Veneto o di alcuni Comuni) che potrebbero adeguare le loro funzioni di gateway al fine di semplificare l'integrazione con i servizi di emissione e verifica delle attestazioni di attributi digitali. Altresì le biblioteche pubbliche potrebbero estendere le loro funzioni di reference desk assumendo il ruolo di facilitatori digitali (informazioni, assistenza operativa ecc.) verso i cittadini nella fase di accompagnamento post rilascio dell'IT-Wallet e dei portafogli accreditati.

La trasformazione è appena avviata, il percorso non è breve e presenta degli ostacoli, ma occorre predisporci per affrontarlo senza precondizionamenti, seguendone le evoluzioni e valutando strada facendo le opportunità offerte da questa innovazione.

³⁸ Il progetto "Rete dei servizi di facilitazione digitale" (Misura 1.7.2 del PNRR) prevede la costituzione di tremila presidi "Digitale facile" per accrescere le competenze e l'inclusione digitale di oltre 2 milioni di cittadini entro il 2026: <https://repubblicadigitale.gov.it/portale/progetti-del-dipartimento/rete-dei-servizi-di-facilitazione-digitale>.

Riferimenti bibliografici*

- BL (British Library). 2024. *Learning lessons from the cyber-attack*. British Library cyber incident review, 8 marzo. [https://www.bl.uk/home/british-library-cyber-incident-review-8-march-2024.pdf/](https://www.bl.uk/home/british-library-cyber-incident-review-8-march-2024.pdf).
- Clusit (Clusit - Associazione Italiana per la Sicurezza Informatica). 2025. *Rapporto Clusit sulla Cybersecurity in Italia e nel mondo 2025*. Astrea. https://clusit.it/wp-content/uploads/download/Rapporto_Clusit_03-2025_web.pdf.
- Codice per l'amministrazione digitale*. D.Lgs. 7 marzo 2005, n. 82 e successive modifiche. <https://www.normattiva.it/uri-res/N2Ls?urn:nir:stato:decreto.legislativo:2005-03-07;82>.
- European Commission. 2024. *CIR 2024/2977, Commission Implementing Regulation (EU) 2024/2977 of 28 November 2024 laying down rules for the application of Regulation (EU) No 910/2014 of the European Parliament and of the Council as regards person identification data and electronic attestations of attributes issued to European Digital Identity Wallets*. https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=OJ:L_202402977.
- European Commission. 2025. *European Digital Identity Wallet Architecture and Reference Framework*. <https://eu-digital-identity-wallet.github.io/eudi-doc-architecture-and-reference-framework/latest/>.
- European Parliament and Council. 2014. *Regulation (EU) No 910/2014 of the European Parliament and of the Council of 23 July 2014 on electronic identification and trust services for electronic transactions in the internal market and repealing Directive 1999/93/EC*. <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:02014R0910-20241018>.
- Grove, Jack. 2024. "How the British Library cyberattack disrupted research." *Times Higher Education*, 21 novembre. <https://www.timeshighereducation.com/depth/how-british-library-cyberattack-disrupted-research>.
- IETF (Internet Engineering Task Force). 2000. *RFC 2798 Definition of the inetOrgPerson LDAP Object Class*. <http://tools.ietf.org/html/rfc2798>.
- IETF (Internet Engineering Task Force). 2006a. *RFC 4519 Lightweight Directory Access Protocol (LDAP): Schema for User Applications*. <http://tools.ietf.org/html/rfc4519>.
- IETF (Internet Engineering Task Force). 2006b. *RFC 4524 COSINE LDAP/X.500 Schema*. <http://tools.ietf.org/html/rfc4524>.
- IETF (Internet Engineering Task Force). 2012. *The OAuth 2.0 Authorization Framework*. Internet Engineering Task Force (IETF) Request for Comments 6749. <https://datatracker.ietf.org/doc/html/rfc6749>.
- Kime, Chad. 2024. *6 Best Enterprise Password Managers for 2024*. *eSecurity Planet*. <https://www.esecurityplanet.com/products/best-password-managers/>.
- OASIS Open. 2008. *Security Assertion Markup Language (SAML) V2.0 Technical Overview*, OASIS Security Services TC Draft 02. <http://docs.oasis-open.org/security/saml/Post2.0/sstc-saml-tech-overview-2.0.html>.

* Ultimo accesso di tutti i riferimenti online, compresi quelli citati nelle note, il 31 marzo 2025.

OIDF (OpenID Foundation). 2023. *OpenID Connect Core 1.0 incorporating errata set 2.* https://openid.net/specs/openid-connect-core-1_0.html.

REFEDS (Research and Education FEDerations). 2022. *SCHAC – SCHema for Academia.* <https://wiki.refeds.org/display/STAN/SCHAC>.

REFEDS (Research and Education FEDerations). 2023. *EduPerson Object Class Specification.* <https://wiki.refeds.org/display/STAN/eduPerson>

W3C (World Wide Web Consortium). 2025. *Verifiable Credentials Data Model v2.0.* W3C Proposed Recommendation. <https://www.w3.org/TR/vc-data-model-2.0/>.

Participatory Democracy and the Washing Machine Syndrome: Lights and Shadows of Inclusive Archiving between Demagogic and Social Participation

Giulia Guidarelli^(a)

a) University of Bologna, <https://orcid.org/0009-0008-4489-3978>

Contact: Giulia Guidarelli, g.guidarelli3@studio.unibo.it

Received: 27 April 2025; Accepted: 17 June 2025; First Published: 15 September 2025

ABSTRACT

This article offers an analysis of the wide spectrum of participatory archival phenomenologies and practices, with the aim of identifying the critical issues inherent in these processes of documentary sedimentation, particularly in relation to their defining feature: the ethical commitments that shape them. It critically examines the role of archivists within frameworks of reparative social justice, drawing attention to the potential risks of instrumentalizing both the “democratization of history” and the use of digital resources. The discussion ultimately addresses archival responsibility toward the emotional and informational needs that give rise to participatory archives, grounding them in a physical and community-based dimension.

KEYWORDS

Community archives; Participatory archives; Living archives; Inclusive archiving.

Democrazia partecipativa e sindrome della lavatrice. Luci e ombre di un’archivistica inclusiva tra demagogia e compartecipazione sociale

ABSTRACT

Il contributo sviluppa un’analisi del ventaglio di fenomenologie e prassi archivistiche partecipative con lo scopo di evidenziare i punti critici di tali processi di sedimentazione documentaria soprattutto alla luce del loro elemento qualificante: il comportamento etico che le anima. L’articolo prende pertanto in considerazione il ruolo degli archivisti rispetto a dinamiche di giustizia sociale riparativa, ponendo l’accento sui rischi dettati da una strumentalizzazione dei processi di “democratizzazione della storia” e delle risorse digitali. La riflessione, infine, verte sulla responsabilità archivistica nei confronti dei bisogni al contempo emotivi e informativi a partire dai quali si sviluppano gli archivi partecipati, ancorandoli a una dimensione fisica e comunitaria.

PAROLE CHIAVE

Archivi di comunità; Archivi partecipativi; Archivi viventi; Archivistica inclusiva.

Introduzione

Il processo di sviluppo e diffusione di fenomenologie archivistiche fortemente orientate verso gli utenti e basate su principi quali partecipazione, giustizia sociale e senso d'appartenenza ha raggiunto ormai uno stadio avanzato, soprattutto nel mondo anglosassone, avendo mosso i primi passi sul finire degli anni '90 con gli *invented archives* (O'Malley e Rosenzweig 1997) e i *community led archives* (Gilliland e Flinn 2014). Tali realtà, poi, hanno recentemente assunto la fisionomia generativa e "rivoluzionaria" dei cosiddetti *living archives* (Almeida e Hoyer 2020). Di fronte a tali fenomenologie, per quanto già immersa in logiche *postcustodial* e intenta a costruire un rapporto più o meno sano con le Information and Communications Technologies (ICT), la disciplina archivistica è costretta a porsi una serie di interrogativi che possano rinnovare il mestiere e la sua consolidata volontà di servizio. Gli esiti di questa (auto)analisi potrebbero consentire alla materia di imboccare definitivamente una strada a onore del vero in parte già battuta (Huivila 2008; Flinn 2011; Bertella Farnetti, Bertuccelli, e Botti 2017; Silveira 2018; Pezzica 2020; Valacchi 2024a), al bivio con discipline come la Public History e l'Information Literacy, con ricadute considerevoli sia sul piano tecnico sia sul piano etico.

Concretamente, non è semplice scindere i due livelli, poiché la documentazione frutto di pratiche partecipative – prodotta, conservata e gestita in linea con un'apertura delle istituzioni tradizionali o a dispetto del loro esplicito dissenso – risulta fortemente connotata da un punto di vista etico. Proprio la difficoltà di ignorare la rivendicazione politica, civile, sociale che vivifica le fenomenologie partecipative del settore spinge la comunità archivistica a scelte professionali che si legano a questioni di metodo e umane al tempo stesso.

Lo scopo di questo articolo è pertanto quello di indagare l'elemento qualificante delle risorse nate in contesti archivistici partecipativi ossia, come si è detto, il comportamento etico che determina e anima la loro produzione. Dopo aver illustrato lo spettro di forme partecipative e aver discusso le varie etichette con cui sono state contrassegnate le realtà archivistiche prese in considerazione, il vero focus della riflessione vuole essere posto sull'agire umano dietro le stesse. La comprensione di tali azioni – anche se in numerosi casi andrebbero piuttosto considerate come reazioni – permetterà di fornire alle risorse una necessaria contestualizzazione. Il fine ultimo dell'articolo sarà, pertanto, considerare la posizione assunta dagli archivisti rispetto ai bisogni emotivi e informativi alla base delle fenomenologie partecipative prese in esame.

1. Lo spettro delle fenomenologie partecipative: le forme

Senza perderci in un lavoro tassonomico apparentemente fine a sé stesso, risulta fondamentale illustrare innanzitutto il ventaglio delle fenomenologie archivistiche partecipative: l'intento è quello di offrire un punto di riferimento per la successiva analisi delle dinamiche sottese all'agire dell'utenza negli archivi. Le categorie che hanno al centro l'idea di partecipazione, pertanto, non sono da interpretare come realtà chiuse e limitanti, dal momento che spesso risultano in comunicazione fra loro, quando non addirittura, in determinati contesti, sovrapponibili, senza dimenticare che il loro habitat naturale è il web che a sua volta asseconda contaminazioni.

Il dialogo che connota le forme che gli archivi assumono attraverso l'adozione di prassi partecipative, abbinato alla natura liquida ed eterogenea di sedimentazioni archivistiche digitali (Valacchi

2020, 83), dà vita a un'ampia gamma di risorse. Queste ultime inducono a ridefinire l'archivio sotto vari aspetti. Da un punto di vista concettuale, infatti, tali aggregazioni scardinano il punto di riferimento dato dal principio di provenienza e tendono a sfuggire dalla definizione canonica di archivio. Oltre a ciò, guardando gli aspetti pratici di produzione, gestione e conservazione documentaria, esse comportano un cambiamento nella fisionomia tanto del soggetto produttore quanto dell'utente. Più precisamente, questo cambiamento consiste in una sincrasì fra i due: la figura del *prosumer*, pur essendo stata definita nel corso degli anni '70 (Toffler 1980), entra nell'ambito archivistico in tempi relativamente recenti (Rosenzweig 2006), legandosi – come appena accennato – a una concezione elastica e virtuale della provenienza tipica del panorama *postcustodial*. “Provenance, in short, is linked to function and activity rather than structure and place” (Cook 2000, 17). Questa ibridazione indotta dall'uso del web e da varie dinamiche della società contemporanea, dunque, produce effetti inevitabili sugli archivi e, per ovvi motivi, sul lavoro degli archivisti, incidendo in particolare sull'attività di mediazione. L'affermarsi di prospettive disciplinari collaborative e l'interesse da parte dei *prosumers* per le testimonianze di *governance*, ossia per quei processi non istituzionali che animano la vita sociale di gruppi e individui, rappresentano la premessa concreta e reale di qualsiasi tipizzazione teorica e astratta si voglia elaborare in merito a fenomenologie archivistiche partecipative.

1.1 Invented archives

On the Web the past is deeply embedded in the present in ways that escape our notice in the conventional archive or library. [...] The Web offers one key departure – it lets users produce their own version of history (O'Malley e Rosenzweig 1997, 134).

Per quanto da un punto di vista complessivo non risulti, naturalmente, *up to date*, O'Malley e Rosenzweig offrono in questo articolo considerazioni ancora valide per iniziare a illustrare una tipologia di archivi in cui è ravvisabile una volontà partecipativa, seppur asistemistica e in un certo senso meno “raffinata” di altre. La commistione di passato e presente che si realizza attraverso la rete, infatti, è stata e continua a essere un'arma a doppio taglio, in grado di innescare ipotesi di democratizzazione della storia che possono avere importanti ricadute sulla produzione, la gestione e la conservazione documentaria.

Gli *invented archives* erano già presenti nel web in forme “embrionali” sul finire degli anni '90, ma è solo con i tragici fatti dell'11 settembre 2001 che il loro sviluppo giunge a compimento (Silveira 2018). Esempio per eccellenza di questo tipo di aggregazioni documentarie è per l'appunto il *September 11 Digital Archive*,¹ posto in essere attraverso una pratica di crowdsourcing svoltasi sotto l'ala di un eminente istituto quale il Roy Rosenzweig Centre For History and New Media (RRCHNM).² Esso presenta tutte le caratteristiche che hanno portato a definire l'archivio inventato come

¹ The September 11 Digital Archive: <https://911digitalarchive.org/about>.

² Roy Rosenzweig Center for History and New Media: <https://rrchnm.org/our-story/>.

un insieme di oggetti digitali di diverso formato e tipologia, generalmente svincolati dai loro soggetti produttori e sottoposti a un assemblaggio fortemente soggettivo ... Siamo di fronte a un processo di riqualificazione concettuale e formale che va oltre alle implicazioni di un approccio per materia, perché non si applica ai contenuti, ma alla stessa ragion d'essere dell'archivio (Valacchi 2024a, 105-6).

La ricontestualizzazione che connota tali aggregazioni documentarie è dettata da una fortissima componente emotiva; in questo senso, la prospettiva storica dell'archivio perde di profondità e si adatta a interessi singolari, profondamente intrecciati con la realtà presente (il “deeply embedded” di cui sopra). Paradigmaticamente, il *September 11 Digital Archive* è sì una testimonianza storica, ma ha un intento patriottico piuttosto evidente, ed è proprio questa matrice che, innestandosi su una ferita collettiva, rende la testimonianza stessa uno strumento che mira a orientare la reazione dell'utente.

Il risultato di questa manipolazione è un'aggregazione documentaria quantitativamente notevole, ma discutibile nella qualità delle scelte metodologiche. Ciò priva gli archivi della loro intrinseca polivalenza e ne limita l'autorevolezza, mettendo in secondo piano i “naturali” processi di sedimentazione, permettendo allo storico di farsi artefice di contesti, cioè delle sue stesse fonti (Valacchi 2023, 158-61). Anche quando l'aggregatore dichiara una presa di coscienza relativamente ai meccanismi di acquisizione e uso delle testimonianze, come nel caso del più recente *Ukraine War Archive*,³ non si può prescindere dalle chiare implicazioni emotive e dall'invenzione in sé, che finisce inevitabilmente col creare “una verità autorevole per diritto di rete” (Valacchi 2024a, 109).

1.2 Community led archives

Se la parola “aggregatore” (*builder*) risulta appropriata per descrivere le dinamiche sotseste agli *invented archives*, si entra più propriamente nel campo dei *prosumers* con i *community led archives*. Anche in questo caso, per non perdersi nella rapidità dei processi che agitano la società e gli archivi che essa stessa produce, si ritiene opportuno fornire alcune coordinate temporali e geografiche. Benché un certo fermento fosse osservabile fin dagli anni ’60, è nel Regno Unito e nell'Irlanda degli anni ’90 che la costruzione di complessi documentari estranei al circuito tradizionale, a opera di comunità marginalizzate, ha raggiunto le dimensioni di un vero e proprio fenomeno politico-culturale, promosso e approfondito dalla comunità archivistica nel passaggio al nuovo millennio.⁴ Nello stesso periodo gli Stati Uniti hanno visto l'evoluzione di sviluppi analoghi (Gilliland e Flinn 2014) e anche in Sud Africa sono stati affrontati problemi connessi all'esercizio di potere insito nelle prassi archivistiche e alle forme di esclusione di cui l'archivio può essere strumento (Mbembe 2002).

Frutto della necessità umana di documentare un'esistenza trascurata o cancellata dall'ambiente istituzionale, i *community led archives* non nascono a uso e consumo della sola realtà che li alimenta (i già citati *prosumers*), ma si caratterizzano per gli intenti militanti. Marcando un solido

³ “Ukraine War Archive is not a public resource. Access to it can be obtained after a background check and authorization”, Ukraine War Archive: <https://ukrainewararchive.org/eng/#about>.

⁴ Community Archives and Heritage Group (CAHG): <https://www.communityarchives.org.uk/content/about/history-and-purpose>.

legame con l'attivismo politico e sociale, si fanno portatori materiali di un senso d'appartenenza identitario – in molti casi è possibile, perciò, parlarne anche in termini di *activist archives* e *social movements archives* (Flinn 2011; Velte 2018; Humbel 2022). Le complesse vicende socioculturali che alimentano programmaticamente questa fenomenologia archivistica si rispecchiano soprattutto nella volontà di non farsi assorbire dal tessuto documentario tradizionale, favorendo invece indipendenza e autonomia: “Many of these collections and independent institutions [...] have eventually made their way into professional, often higher education, repositories but others have remained largely independent and autonomous, committed to retaining control over the resources for the production of their history” (Gilliland e Flinn 2014, 6).

1.3 Participatory archives

Indipendenza e autonomia sono da intendere come sinonimi di orgoglio e rivendicazione, non di isolamento. La ricchezza dei *community led archives*, al di là dell'insieme di memorie collettive che senza la loro esistenza rischierebbe l'oblio, sta in ciò che essi costruiscono: un dialogo animato da voci diverse ma accomunate dall'essere rimaste inascoltate, e quindi bisognose di amplificarsi reciprocamente. Proprio questa dinamica costruttiva è l'elemento fondante degli archivi partecipativi, nei quali l'azione di realtà diverse dai consueti agenti istituzionali incontra innanzitutto le necessità dei propri simili, ma finisce poi con l'influenzare anche l'ambiente archivistico più tradizionale.

Tali sviluppi sono certamente esito delle opportunità interattive offerte dal web 2.0, ma hanno subito un'accelerazione significativa nell'ultimo quinquennio. Le restrizioni dovute alla pandemia di COVID-19, difatti, hanno imposto agli istituti della cultura non solo di ingegnarsi per essere raggiungibili, ma soprattutto di essere loro stessi a raggiungere per primi gli utenti. Non senza un pizzico di ironia, è stata la chiusura imposta dal distanziamento sociale a spingere davvero le realtà conservative canonicamente intese verso un'apertura del loro punto di vista nei confronti del panorama documentario fino a quel momento ritenuto estraneo. D'altronde, sono state anche le stesse comunità, nel sommarsi dei problemi dovuti all'isolamento alle difficoltà preesistenti, a cercare forme di contatto nuove e diverse anche attraverso i prodotti culturali. Si vedano per esempio il progetto GLAMMONS⁵ e il progetto #NorthVanStories – Living History.⁶ Quest'ultimo è stato illustrato dettagliatamente da Bushey (2023).

Si arriva quindi a parlare di archivi 2.0 *user centered* (Theimer 2011) e di archivi “of the people, by the people, for the people” (Evans 2007). L'utente, con tutta la sua umanità, entra nel merito del processo di produzione, conservazione e gestione documentale negoziando con le istituzioni. Tale negoziazione si lega contemporaneamente a principi e pratiche condivise dalle comunità (Gilliland e McKemmish 2014) e a questioni di autorialità, privacy e finalità dell'archivio (Eveleigh e Benoit 2019). Per poter riunire questi due mondi – come si approfondirà in seguito – l'archivista indossa i panni che dovrebbero essergli più consoni, ovvero quelli di mediatore culturale. La disciplina archivistica “al fondo rimane una scienza di trasmissione e comunicazione” (Valacchi 2024a, 48) e

⁵ GLAMMONS: <https://glammons.eu/>.

⁶ #NorthVanStories – Living History: <https://monova.ca/northvanstories/>.

viene traghettata dalle acque tutto sommato placide di una dimensione esclusivamente filologica, storica, giuridica a quelle impetuose del dibattito sull'umanesimo digitale, della diffusione della conoscenza e della costruzione di una storia pubblica in quanto partecipata.

Abbiamo già notato come le risorse digitali sembrano svolgere in questa “rotta” partecipativa la funzione di un faro: un riferimento per i timonieri (gli archivisti) e i loro passeggeri (gli utenti), eretto dagli abitanti della terraferma (le istituzioni) per garantire al viaggio una lieta conclusione. Ora, non sembra opportuno in questa sede entrare nel merito dei possibili malfunzionamenti di un faro, ma fuor di metafora è possibile dire che il ruolo guida attribuito alle risorse digitali spesso non ha trovato riscontro nella realtà. L'utilizzo del web per realizzare prassi partecipative ha ricadute notevoli: “notions of co-creation and parallel or simultaneous multiple provenance reposition ‘records subjects’ as ‘records agents’” (Gilliland e McKemmish 2014, 5). Di conseguenza, per rispettare quanto elaborato nella più ampia agenda di giustizia sociale, trasformazione politica, rivendicazione di diritti civili cui fanno capo, tali prassi devono assumere nel loro lato più strettamente archivistico la forma di un processo ampio e lungimirante di alfabetizzazione digitale (Valacchi 2024a, 117).

1.4 Living archives

Senza anacronismi, tuttavia, c’è uno spazio diverso rispetto a quello virtuale che non va dimenticato: una dimensione fisica da riscoprire nelle sue potenzialità, certamente anche aumentate, purché all’archivio sia riconosciuta la dinamicità trasformativa che gli appartiene e che non dipende dall’uso delle risorse digitali. L’attività svolta dai *prosumers*, d’altronde, pur risultando evidente sul web, ha modo di essere visibile anche offline, anzi merita particolare attenzione proprio quando dà vita a fenomenologie che si nutrono di socialità e comunità, che nascono per essere vissute e per essere esaltate nella loro dimensione corrente.

By collecting diverse narratives that can be “deconstructed and reconstructed” archives can re-activate social and political agency formerly stripped away by capitalism, and can provoke critical reflection on tools used in emancipatory struggles in order to reclaim individual and collective memory and create opportunities for performative action that enables new socio-economic possibilities (Almeida e Hoyer 2020, 14).

I *living archives* si pongono obiettivi di cambiamento, secondo una prospettiva generativa e performativa già vista nei *community led archives*. Questo spirito battagliero è il portato diretto dello sviluppo di una coscienza culturale attiva all’interno di comunità poco o affatto rappresentate negli archivi istituzionali. Esso crea innanzitutto una maggiore interazione all’interno della comunità stessa, ampliando le basi partecipative dell’archivio; il livello successivo di questa estensione dell’*agency* culturale è il coinvolgimento – emotivo e informativo – della società tutta verso l’archivio e chi lo costruisce, stimolato attraverso attività concrete; infine, ottenuto un supporto collettivo, lo scopo ultimo è riuscire a trasformare le linee governative che hanno escluso la comunità dal circuito della conservazione documentale. Questo movimento dal basso, una vera e propria “lotta informativa” (Valacchi 2024a, 120), è ben rappresentato nel seguente schema (Lian e Olivier 2018):

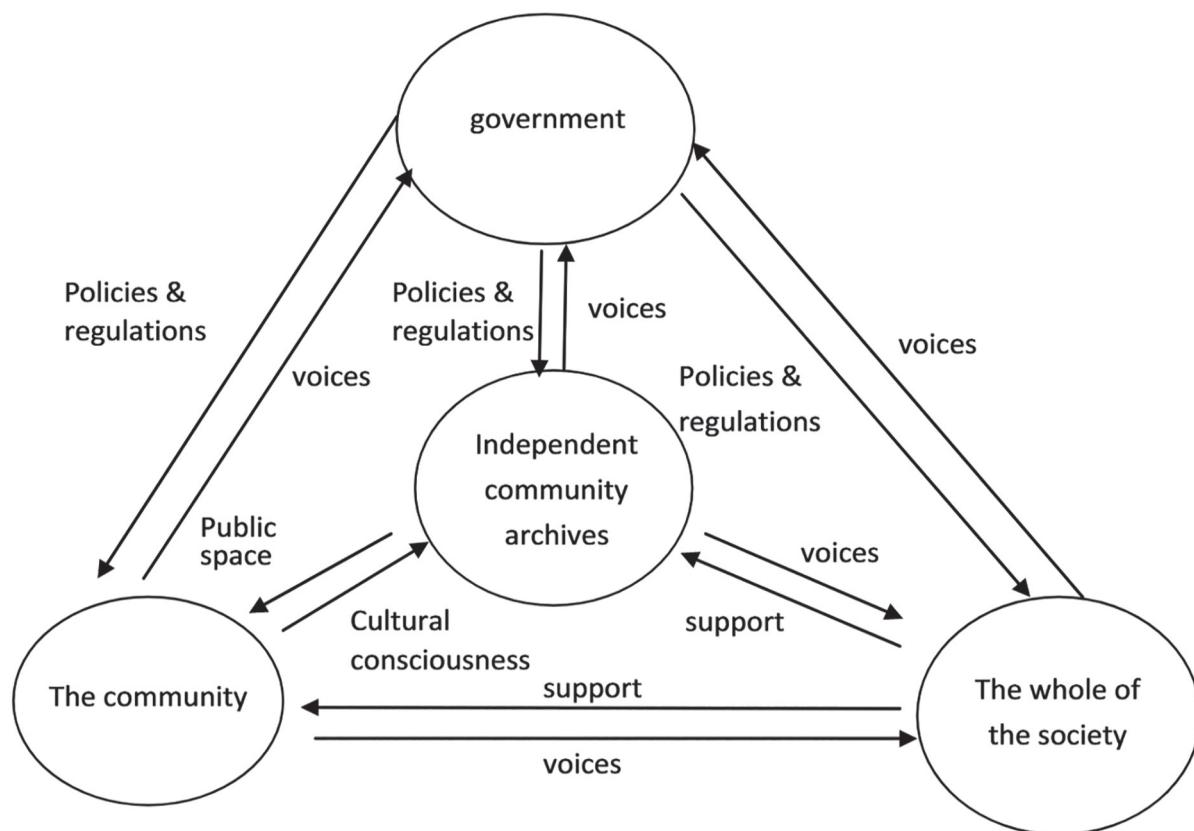


Figura 1. Lo schema illustra le dinamiche che dovrebbero regolare i rapporti tra *community archive*, comunità che lo origina, autorità governativa di competenza e società nel suo insieme; mostra anche come i quattro ambiti possano influenzarsi reciprocamente.

Pur essendo stato realizzato con lo scopo di sintetizzare le sfide affrontate dai *community led archives* in Cina, lo schema risulta comunque adeguato per illustrare le analoghe dinamiche che i *living archives* puntano a mettere in moto nel mondo occidentale.

2. Lo spettro delle prassi partecipative: i meccanismi

Alla luce della panoramica appena conclusa, è possibile affermare che le fenomenologie archivistiche partecipative esaltano il potenziale che l'archivio ha sempre avuto in quanto “presupposto ineludibile per l'attivazione di qualsivoglia processo di *capability*, ovvero di accesso del singolo alle possibilità di sviluppo individuale, in termini di opportunità (somma di beni e servizi e relazioni)” (Rossi 2021, 47).

Si parla, dunque, di welfare: l'archivio come strumento di *capability* si rafforza attraverso le prassi partecipative per tutelare i diritti civili. Di conseguenza, allo spettro delle fenomenologie partecipative fa da corrispettivo uno spettro di criteri che regolano l'attività dei *prosumers*, strutturando il rapporto tra archivio, comunità che lo vivificano e istituzioni. Le riflessioni della *project archivist* Hannah Mackay (2019) sulla definizione di archivio partecipativo, confluite nel seguente grafico [Figura 2], e le sue considerazioni in merito alle criticità metodologiche insite

in queste fenomenologie rappresentano un punto di partenza essenziale per definire il campo d'azione dei *prosumers*.

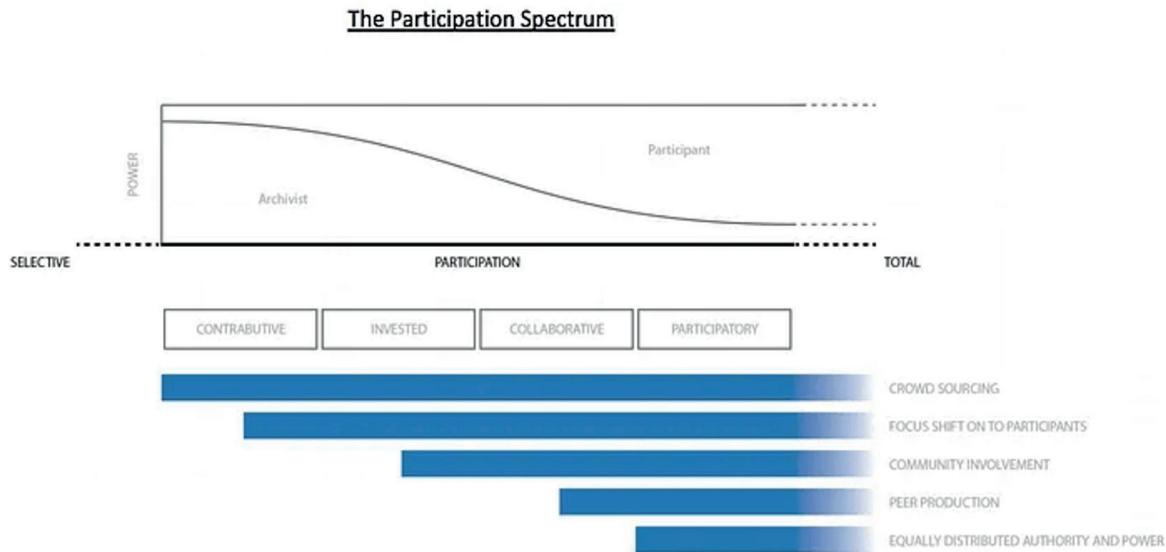


Figura 2. Il grafico propone un ventaglio delle prassi archivistiche partecipative.

2.1 *Crowdsourcing*: per l'istituzione

Prescindendo da una partecipazione passiva, ossia dalla semplice attività di consultazione, il crowdsourcing è il primo mezzo per l'attuazione effettiva di prassi partecipative in ambito archivistico. Attraverso la cosiddetta “*call for documents*”, gli archivi tradizionali acquisiscono documentazione (inclusi dati social) dalla cittadinanza per costruire nuovi fili gnoseologici (Velte 2018). Tuttavia, benché sia una pratica amata soprattutto dalla Public History (Pezzica 2020, 99-105), il crowdsourcing si rivela in maniera evidente come una forma di partecipazione orientata quasi completamente verso l'istituzione: il circuito tradizionale della gestione e conservazione documentale apre le sue porte al mondo esterno, continuando però a favorire una struttura gerarchica e piramidale, in cui l'azione dell'utente è limitata a un approccio contributivo.

2.2 Dal contributo saltuario alla collaborazione impegnata: per il pubblico

Ciò che conferisce maggiore incisività all'utente e lo trasforma gradualmente in *prosumer* è la possibilità di collaborare attivamente con l'istituzione. Perché ciò sia possibile, quest'ultima ha il dovere di intercettare i bisogni informativi della cittadinanza.

Gli effetti di questo slancio, abbinato spesso all'impiego delle funzionalità interattive del web 2.0 e quindi alla creazione di archivi ibridi, hanno ricadute molto pratiche che possono intimorire gli archivisti: come anticipato, i contesti di produzione documentale collaborativa impongono di rivalutare la fiducia finora riposta nel principio di provenienza poiché generano problemi di accuratezza e autenticità. Quando l'archivio viene orientato senza soluzione di continuità dall'utente e verso l'utente, difatti, il rischio concreto è dato dall'impossibilità di ricostruire la filiera documentale.

La perdita di informazioni di contesto, anche qualora riguardi una singola unità documentaria, può ledere l'autorevolezza complessiva dell'archivio. È possibile ovviare a tale problema non solo attraverso la scelta dei giusti strumenti informatici e digitali, ma soprattutto dando peso all'*iter* del lavoro archivistico, invece che ai suoi prodotti: l'approccio è sintetizzabile nella formula “iterative products, not ‘perfect’ products” (Theimer 2011, 63). Inoltre, abbandonando per un momento il lato teorico, vale la pena ricordare che, nel concreto, analoghe difficoltà affliggono da sempre anche gli archivi tradizionali; spia evidente di dinamiche non proprio trasparenti, per esempio, è la diffornità diffusa tra scarto legale e scarto reale. Il calibro delle risorse archivistiche collaborative va misurato allora non tanto in base alla loro perfetta accuratezza, quanto semmai tenendo in considerazione il fatto che esse costituiscono testimonianze capaci di arricchire quadri informativi incompleti. I contributi fungono da “important supplements, counters, or corrections to the records held in official archives” (Gilliland e McKemmish 2014, 80): ecco ciò che conferisce loro dignità. Il processo collaborativo, quando indirizzato verso l'utente più che verso l'istituzione, porta in primo piano anche la complicata questione dell'accesso. La domanda è la seguente: come trattare le informazioni sensibili salvaguardando contemporaneamente gli interessi dell'istituzione e della comunità da cui deriva il bisogno informativo intercettato, senza trascurare le eventuali differenze antropologico-culturali che li separano? Per sottrarsi al “rischio di etnocentrismo” (Valacchi 2024a, 113), la risposta prevede la negoziazione del consenso: condividere responsabilità conservative significa interessarsi genuinamente al benessere di tutte le parti coinvolte circa l'esito della collaborazione – nonostante l'impossibilità ontologica di prevedere tutti gli usi futuri di un archivio. Perché tali condizioni siano soddisfatte, il consenso deve avere precise caratteristiche: deve essere preventivo, ossia negoziato preliminarmente; libero, quindi estraneo a condizionamenti esterni; informato, il che si traduce per l'istituzione e per gli archivisti in un obbligo di divulgazione chiara e trasparente dei termini della collaborazione (Gilliland e McKemmish 2014, 84-5). Tali requisiti garantiscono ai *prosumers* una progettualità effettivamente inclusiva e una conservazione efficacemente tutelata anche rispetto alle ingerenze governative che possono verificarsi in contesti di oppression.

2.3 Gerarchie infrante: partecipazione attiva

L'assetto appena illustrato si basa indiscutibilmente sulla disponibilità da parte delle istituzioni a fare dei passi indietro, concedendo un nuovo margine di autorità agli utenti; tuttavia, non altera la struttura gerarchica del circuito di produzione, gestione e conservazione documentaria.

Non che l'archivio partecipato miri a esautorare l'archivista. Al contrario, riconosce la necessità di una figura competente di riferimento. In contesti di autorialità condivisa, difatti, da un lato occorre disporre di un bagaglio professionale adeguato, così da individuare in base a requisiti archivistici minimi la documentazione prodotta e raccolta (Huvila 2008, 20), dall'altro è doveroso riconoscere pari autorevolezza ai bisogni dell'utente e a tali competenze. In questo senso, l'attività di mediazione produce attendibilità e fiducia solo se e quando evolve nei termini di una disseminazione archivistica. Militante è l'archivio, militante è l'archivista (Valacchi 2021, 70): la disseminazione scardina le dinamiche *top-down* e ricostruisce l'ambiente archivistico come un ambiente a tutti gli effetti cooperativo. Oltre a dettare le linee guida di progetti, la cui trasparenza può

aiutare ad attrarre l'utente e consolidarne l'*agency* (Mackay 2019), l'archivista asseconda l'indirizzo comunicativo della disciplina per condividere valori etici (Valacchi 2021, 83-4). La mediazione, quando diventa disseminazione, non è più rivolta al singolo, ma si predisponde ad accogliere una molteplicità di pubblici (Valacchi 2021, 85).

Accanto alle competenze oggettive, che sono fonte di sicurezza e quindi garantiscono un'attrazione verso l'archivio, anche la sensibilità con cui sono accolte le istanze dei *prosumers* alimenta un senso d'appartenenza. La comprensione e la condivisione della matrice etica delle fenomenologie archivistiche partecipative sfociano in un atto di democrazia che avvalora i progetti di *shared authority*. Alla luce di questa dimensione civica, come dimostrano perfettamente i *living archives*, l'archivio diventa “luogo immediato dove fare politica” (Valacchi 2024a, 120).

Ciò lega gli archivisti alla metodica protesta che dà vita a sedimentazioni e aggregazioni documentarie partecipate. Per la sopravvivenza stessa della disciplina, “dedication to the past is not enough” (Ketelaar 1997): senza perdere contatto con la realtà presente, considerando criticamente le ricadute future dell'attività di disseminazione, la comunità archivistica si predisponde ad accogliere bisogni al contempo emotivi e informativi. Ma affinché gli archivi partecipativi mantengano la loro attendibilità e rappresentatività una volta adottata tale prospettiva, sarà opportuno tutelarli contro eventuali sentimentalismi e strumentalizzazioni, evitando di cadere nelle trappole del cosiddetto “*white saviour complex*” e di una deificazione delle risorse digitali.

3. Lo spettro dei comportamenti partecipativi: gli aspetti etici

Sfatare i miti che sono stati via via costruiti attorno alle fenomenologie archivistiche partecipative è una questione di credibilità e onestà intellettuale: significa, cioè, fare ritorno a una salda deontologia professionale per gestire la complessità di precise dinamiche. Difatti, la narrazione di una riuscita democratizzazione della storia è sì lusinghiera per le stesse istituzioni e in apparenza positiva per chi ne beneficia, ma a tratti si rivela piuttosto stridente. In maniera analoga, come già anticipato, la gran fiducia riposta nel salvifico mondo digitale manca di una visione globale che sappia trovare risposte più esaustive a un'evidente sofferenza informativa.

3.1 Democratizzazione della storia: una narrazione

Il motto “We use digital media to democratize history” del già citato RRCHNM è un ammaliante canto di sirene da affrontare non con i tappi di cera, ma con atteggiamento critico (Valacchi 2024b, 57). A dover essere oggetto d'attenzione è proprio la bella parola “democrazia”, e il modo in cui il suo significato viene circoscritto dalle istituzioni. Le forme e le prassi attraverso le quali si attua la partecipazione in ambito archivistico, in base a quanto esplorato precedentemente in questo contesto, rispecchiano per l'appunto diverse concezioni di un simile processo di democratizzazione. È possibile osservare come si parta da una visione più distorta, passando poi per prospettive effettivamente inclusive, fino a raggiungere un desiderio di eversione rispetto alla preservazione e tradizione della memoria documentale.

La volontà di introdurre *community archives* e *participatory archives* all'interno dei consueti circuiti di conservazione e fruizione documentaria, più che ad autentici intenti di giustizia sociale riparativa

fa capo non di rado al bisogno di un ritorno d'immagine. Proprio come avviene per fenomeni quali *greenwashing* e *pinkwashing*, allo stesso modo i fondi documentari che nascono con l'intento di dare voce alle *underdocumented populations* corrono il rischio di finire nella lavatrice istituzionale: si avvia pure il programma rapido antimacchia, per dare l'idea di aver rimosso l'onta di una complicità nella marginalizzazione delle comunità coinvolte. In quest'ottica, la democratizzazione della storia si affina come strategia comunicativa possibilmente vicina al *wokewashing*⁷, imponendo l'asservimento delle fenomenologie partecipative a una narrazione che in realtà ha ben poco di democratico e nasconde semplicemente altra oppressione. Il gruppo nordamericano Faculty Organizing for Community Archives Support (FOCAS) parte da una prospettiva accademica per mettere in evidenza, tra le altre, anche tali difficoltà: per formare studenti in grado di comprendere le dinamiche di ambienti archivistici partecipativi e di inserirsi al loro interno con rispetto e flessibilità è necessario non solo arricchire i programmi didattici, ma anche promuovere inclusività negli stessi corsi di studio, cioè garantire un protagonismo professionale alle comunità interessate (Caswell et al. 2024). In quest'ottica, i *living archives* sfuggono totalmente al lavaggio istituzionale, salvaguardati da una generatività effettivamente rivoluzionaria e antitetica rispetto alle realtà conservative tradizionali.

3.2 Democratizzazione della storia: una contronarrazione

Si è già accennato al fatto che la proprietà principale dei *living archives*, ossia la loro fisicità, possa sembrare un ostacolo rispetto alla spinta data dal web 2.0 all'ingresso delle pratiche partecipative in ambito archivistico. Quel "ricordo narrativo" (Valacchi 2024b, 52) che l'archivio digitale è già di per sé, a fronte di una molteplicità di produttori altrimenti impensabile, si va a comporre di innumerevoli fili di memoria. Ma, ribadiamo, le potenzialità interattive non sono prerogativa delle risorse offerte dal web: l'archivio è *hic et nunc* – lo è sempre stato –, ed è così perché tale lo creano gli individui che lo vivono. Grazie a questa immanenza ha una posizione privilegiata in meccanismi identitari che, tenendo d'occhio il passato, animano il presente e possono quindi avere un impatto sul futuro. Quando l'oppressione è sistemica e sistematica, la prospettiva cooperativa può apparire inaccettabile; è in simili casi che uno spazio condiviso come quello del web diventa non priorità ma supporto di una realtà archivistica viva e quotidiana, in evoluzione e in rivolta. Un rifiuto così netto della trappola della "democratizzazione della storia" merita allora una riflessione nuova sull'impatto dell'archivio colto nella sua dimensione spaziale:

For groups who have been symbolically annihilated by mainstream media and archives, community controlled spaces are symbols of autonomy and authority to create complex forms of self-representation. Such physical spaces resonate on a deeply affective level for community members and are both personal and political sites (Caswell et al. 2018, 89).

L'archivio "dal vivo" diventa un simbolo, prova tangibile delle radici passate di una comunità, ancora per il presente; è una casa, un luogo accogliente dove trovare rifugio e coltivare il dialogo tra generazioni di ieri e di oggi; un quartier generale dove sviluppare strategie sociopolitiche e cul-

⁷ <https://digitalstrategyconsulting.com/content/content-marketing/cannes-lions-unilever-boss-warns-woke-washing-hampering-brand-purpose/30842/>.

turali rivolte alla sopravvivenza e al benessere comune (Caswell et al. 2018). Tali scenari inducono a ipotizzare una rivalutazione dell’archivio come “terzo luogo”, concetto sociologico elaborato nel libro *The Great Good Place* di Ray Oldenburg (1989), che definisce uno spazio alternativo all’ambiente casalingo e all’ambiente lavorativo dove l’essere umano possa coltivare interazioni sociali (Pezzica 2020, 45).

E l’archivista? Dove si colloca in tutto ciò, che ruolo assume, quali domande si pone? Partendo dalla consapevolezza che “gli archivi non possono vivere senza interlocutori” (Locatelli 2021, 67), c’è bisogno di una risolutezza ferrea per affrontare il disagio informativo che deriva da intuizioni intime e al tempo stesso comuni, radicali, frutto del rifiuto di un cronico disinteresse a realtà desiderose, grazie allo spazio appunto conquistato, di sentire e farsi sentire. Qui l’archivista si trova di fronte a un’opportunità da non perdere, quella di mettere in campo le proprie competenze come cassa di risonanza: un megafono sia professionale che umano, che sappia dimostrare che l’archivistica è viva e vitale, spettatrice attenta e insieme protagonista colma di curiosità, empatia, iniziativa. La comunità archivistica si predispone a un lavoro “slow and relational, necessitating dispensing with assumptions and listening to tough answers” (Caswell et al. 2024, 403): qualcosa, in sintesi, che sia in grado di decostruire la narrazione della “democratizzazione della storia” per consolidare, invece, processi attivi di compartecipazione sociale.

Conclusione

La nostra riflessione si è concentrata sul comportamento etico che determina e anima la produzione di risorse archivistiche partecipative, tema di fondo tenuto in considerazione prima valutando le forme assunte da tali realtà e poi riflettendo sulle conseguenti questioni metodologiche. L’ultima parte si è poi soffermata sui problemi legati a tali fenomenologie e alle prassi consequenti, nell’intento di focalizzare l’attenzione sulle possibili prospettive per il futuro del mestiere archivistico in rapporto a simili scenari.

Alla luce di quanto osservato, è possibile affermare che la via di un’assunzione di responsabilità da parte del mondo archivistico passa *in primis* attraverso il riconoscimento di una produzione documentaria da parte di comunità esterne al circuito tradizionale e istituzionale, finora altrimenti trascurate o escluse dai processi di cura della memoria. In secondo luogo, proprio in virtù della forte componente etica che accompagna tali aggregazioni documentarie, la disciplina archivistica si ritrova di fronte alla necessità di comprendere e studiare le peculiarità di tali scenari informativi, recependone quando possibile nuove conoscenze e nuovi approcci che integrino il bagaglio professionale e culturale, senza limitare il campo d’azione all’uso di strumenti digitali. Fondamentale, inoltre, appare l’apertura della didattica verso tali ambienti.

Come sottolinea Caswell, gli archivi partecipati hanno un impatto ontologico, epistemologico e sociale dal punto di vista delle comunità che li alimentano, soprattutto in relazione alla loro collocazione spaziale:

ontological impact (in which members of marginalized communities get confirmation “I am here”); epistemological impact (in which members of marginalized communities get confirmation “we were here”); and social impact (in which members of marginalized communities get confirmation “we belong here”) (Caswell 2018, 76).

Certificare che “Io sono in questo posto, Noi siamo stati in questo posto, Noi apparteniamo a questo posto”, significa garantire alla cittadinanza – nella sua varietà propria della realtà contemporanea – di riconoscersi. Significa, cioè, dare modo alle comunità di rintracciare, mantenere e consolidare la propria presenza in un determinato contesto geografico, storico e culturale, dando spessore tangibile all’idea stessa di identità che è valore integrante e centrale della missione archivistica. Qualunque forma abbia l’archivio.

Riferimenti bibliografici*

- Almeida, Nora, e Jen Hoyer. 2020. "Living Archive in the Anthropocene." *Journal of Critical Library and Information Studies* 3 (1). <https://doi.org/10.24242/jclis.v3i1.96>.
- Benoit, Edward III, e Alexandra Eveleigh. 2019. *Participatory Archives: Theory and Practice*. Facet. <https://doi.org/10.29085/9781783303588>.
- Bertella Farnetti, Paolo, Lorenzo Bertucelli, e Alfonso Botti, a c. di. 2017. *Public history. Discussioni e pratiche*. Milano: Mimesis Edizioni.
- Bushey, Jessica. 2023. "A Participatory Archives Approach to Fostering Connectivity, Increasing Empathy, and Building Resilience during the COVID-19 Pandemic." *Heritage* 6 (3): 2379-2393. <https://doi.org/10.3390/heritage6030125>.
- Caswell, Michelle, Sumayya Ahmed, Gracen Brilmyer, Marika Cifor, Jennifer Douglas, Jamie Ann Lee, James Lowry, Vanessa Reyes, Cecilia Salvatore, e Tonia Sutherland. 2024. "Envisioning a Paid Community Archives Internship Program: Challenges and Opportunities." *Journal of Education for Library and Information Science* 65 (4). <https://doi.org/10.3138/jelis-2023-0032>.
- Caswell, Michelle, Joyce Gabiola, Jimmy Zavala, Gracen Brilmyer, e Marika Cifor. 2018. "Imagining transformative spaces: the personal-political sites of community archives." *Archival Science* 18: 73-93. <https://doi.org/10.1007/s10502-018-9286-7>.
- Cook, Terry. 2001. "Archival science and postmodernism: new formulations for old concepts." *Archival Science* 1 (1): 3-24. <https://doi.org/10.1007/BF02435636>.
- Digital Strategy Consulting. 2019. *Cannes Lions: Unilever Boss Warns "woke-Washing" Hampering Brand Purpose*. <https://digitalstrategyconsulting.com/content/content-marketing/cannes-lions-unilever-boss-warns-woke-washing-hampering-brand-purpose/30842/>.
- Evans, Max J. 2007. "Archives of the People, by the People, for the People." *The American Archivist* 70 (2): 387-400. <https://www.jstor.org/stable/40294576>.
- Flinn, Andrew. 2011. "Archival Activism: Independent and Community-Led Archives, Radical Public History and the Heritage Professions." *InterActions: UCLA Journal of Education and Information Studies* 7 (2). <https://doi.org/10.5070/D472000699>.
- Gilliland, Anne, e Andrew Flinn. 2014. "Community Archives: What Are We Really Talking About?" (Keynote, CIRN Prato Community Informatics Conference, 2013).
- Humbel, Marco. "The Digitisation and Open Access Politics of Social Movement Archives." Tesi di dottorato, University of London, 2022. <https://www.proquest.com/docview/2867902292/abstract/7324B06DBAB84CC0PQ/1?sourceType=Dissertations%20&%20Theses>.
- Huvila, Isto. 2008. "Participatory Archive: Towards Decentralised Curation, Radical User Orientation, and Broader Contextualisation of Records Management." *Archival Science* 8 (1): 15-36. <https://doi.org/10.1007/s10502-008-9071-0>.

* Ultima data di consultazione dei siti web: 10 maggio 2025.

Ketelaar, Eric. 1997. "Archives of the People, by the People, for the People." In *The Archival Image. Collected essays*. Hilversum: Uitgeverij Verloren.

Lian, Zhiying, e Gillian Oliver. 2018. "Sustainability of independent community archives in China: a case study." *Archival Science* 18: 313-332. <https://doi.org/10.1007/s10502-018-9297-4>.

Locatelli, Gabriele. 2021. "Diventare visibili. Quattro tempi per una maieutica archivistica." In *Dimensioni archivistiche. Una piramide rovesciata*, a cura di Lorenzo Pezzica e Federico Valacchi. Milano: Editrice Bibliografica: 67-82.

Mackay, Hannah. 2019. "The Participatory Archive: Designing a spectrum for participation and a new definition of the participatory archive." *Medium*. <https://medium.com/@mackayhjc/the-participatory-archive-designing-a-spectrum-for-participation-and-a-new-definition-of-the-964bc1b0f987>.

Mbembe, Achille. 2002. "The Power of the Archive and its Limits." In *Refiguring The Archive*, a cura di Carolyn Hamilton, Verne Harris, Jane Taylor, Michele Pickover, Graeme Reid, e Razia Saleh. Dordrecht: Springer: 19-27. https://doi.org/10.1007/978-94-010-0570-8_2.

Oldenburg, Ray. 1999. *The Great Good Place*. New York: Marlowe & Co.

O'Malley, Michael, e Roy Rosenzweig. 1997. "Brave New World or Blind Alley? American History on the World Wide Web." *The Journal of American History* 84 (1): 132-155. <https://www.proquest.com/docview/224915973/abstract/D0C4A59231604389PQ/9>.

Pezzica, Lorenzo. 2020. *L'archivio liberato. Guida teorico-pratica ai fondi storici del Novecento*. Milano: Editrice Bibliografica.

Rosenzweig, Roy. 2006. "Can History Be Open Source? Wikipedia and the Future of the Past." *The Journal of American History* 93 (1): 117-146. <https://doi.org/10.2307/4486062>.

Rossi, Annalisa. 2021. "La palla di vetro. Immaginare la tutela e la comunicazione archivistica diffusa." In *Dimensioni archivistiche. Una piramide rovesciata*, a cura di Lorenzo Pezzica e Federico Valacchi. Milano: Editrice Bibliografica: 37-62.

Silveira, Pedro Telles da Silveira. 2018. "From Instant History to the Infinite Archive: Digital Archiving, Memory and the Practical Past at the Roy Rosenzweig Center for History and New Media". *Historein* 17 (2). <https://doi.org/10.12681/historein.10964>.

Theimer, Kate. 2011. "What Is the Meaning of Archives 2.0?." *The American Archivist* 74 (1): 58-68. <https://www.jstor.org/stable/23079001>.

Toffler, Alvin. 1980. *The Third Wave*. New York: William Morrow.

Valacchi, Federico. 2020. "Quiddam divinum". Riflessioni sul metodo storico." *Archivi* XV/1.

Valacchi, Federico. 2023. "Se l'archivio è artificiale. Verso uno ius archivi partecipativo?." *AIDAinformazioni* 41 (1): 153-170. <https://doi.org/DOI 10.57574/596529288>.

Valacchi, Federico. 2024a. *L'archivio aumentato. Tempi e modi di una digitalizzazione critica*. Milano: Editrice Bibliografica.

Valacchi, Federico. 2024b. "The Season of Living Archives: A Generative Provenance." *JLIS.it* 15 (2): 51-60. <https://doi.org/10.36253/jlis.it-597>.

Velte, Ashlyn. 2018. "Ethical Challenges and Current Practices in Activist Social Media Archives." *The American Archivist* 81 (1): 112-134. <https://www.jstor.org/stable/48618004>.

The Role of the Information and Documentation Professional in Big Data Management for Scientific Research

Eder Ávila^(a)
Andreu Sulé^(b)

a) Institute for Library and Information Research. UNAM, <https://orcid.org/0000-0001-5439-0495>
b) University of Barcelona, <https://orcid.org/0000-0002-2467-3678>

Contact: Eder Ávila, eder@iibi.unam.mx; Andreu Sulé, sule@ub.edu

Received: 04 February 2025; Accepted: 22 July 2025; First Published: 15 September 2025

ABSTRACT

The role of Information and Documentation professionals (IDPs) in managing big data within the context of scientific research is analyzed. Big data has transformed research by offering new opportunities, such as accelerating results and fostering interdisciplinary fields. However, it also presents significant challenges related to data access, quality, privacy, and security. Information professionals contribute to various areas: data retrieval, organization, description through metadata, and long-term preservation. These multifaceted contributions demand a sophisticated skill set that require not only technical skills (e.g., database management and programming languages) but also interpersonal and ethical competencies that are essential for working in multidisciplinary teams. However, it should not be assumed that all Information and Documentation professionals inherently possess these qualities; rather, facing such challenges demands a continuous effort to cultivate intellectual openness, collaborative practices, and ethical responsibility. By fostering an open data culture and supporting responsible management practices through international standards for interoperability and reuse, Information and Documentation experts play a key role in adding value, structure, and ethics to scientific research in the digital age.

KEYWORDS

Big data; Documentation; Scientific research; Information professional.



Introduction

The nature of big data and its exponential growth has generated — and continues to generate — organizational and management needs in scientific research that have a direct impact on Information and Documentation professionals. Although the omnipresence of digital data and advanced information technology may have given a bigger role to professionals working in the fields of engineering, computer science, and statistics (Almeida et al. 2015), information and documentation professionals have a place on multidisciplinary teams working with big data on academic research projects. Tasks where information and documentation professionals can contribute their knowledge and skills to the development of data-driven research include: searching for and retrieving the most relevant data for each study; cleaning, transforming, and reconciling that data to enable more precise analysis; using or developing ontologies that allow machines not only to process data but also to interpret it and interconnect it with other datasets; and correctly and accurately describing data collections produced by the research center using consolidated metadata standards.

The aim of this article is to explore these and other tasks specific to Information and Documentation professionals working with the big data used on academic research projects. To this end, it begins with a definition of big data, to establish a conceptual framework and define the terminology used. This is followed by a discussion of the opportunities and challenges that big data poses for academic research, emphasizing some of the considerations and precautions to keep in mind when working with this type of data. Finally, an outline is offered of the different tasks where information and documentation professionals can add value to data-driven academic research.

Methodology

A literature review was conducted with the aim of identifying and analyzing information resources that address the theoretical foundations related to the role of information and documentation professionals in big data management within the context of scientific research. To achieve this objective, scientific articles, book chapters, and case studies were reviewed. The resources were obtained from prestigious academic databases such as Elsevier, SpringerLink, IEEE, and Emerald, as well as from institutional repositories like that of National Autonomous University of Mexico (UNAM) and specialized websites, including institutional portals of the European Union and UNESCO. This methodology ensured a comprehensive and rigorous approach to the analysis of the proposed research topic.

The search strategies within these sources focused on the following terms: big data management, scientific research, and information and documentation professionals. It is important to note that, although the selected terms allowed for the retrieval of resources addressing these topics in their titles, a detailed review of each resource was necessary to identify its disciplinary focus. This is because the conceptual notion of the information and documentation professional, as well as the activities related to big data management within the context of scientific research, encompasses multiple study variables, ranging from technical and pragmatic aspects to epistemological considerations.

Therefore, resources published between 2016 and 2024 were selected, taking into account technological and chronological criteria, as the phenomenon under study is constantly evolving due to the increasing number of publications analyzing its impact across various contexts and disciplines. Resources published outside this period were chosen based on their level of contribution and relevance to the analysis of the research topic proposed in this study. This approach ensured that the literature review was up-to-date and representative of the most recent trends and advancements in the field.

In addition to the above, additional filters were implemented to guarantee the quality and relevance of the selected resources. Priority was given to documents that presented an interdisciplinary approach, as big data management in scientific research requires the integration of knowledge from areas such as data science, computer science, library science, and knowledge management. Similarly, special attention was paid to studies offering critical or innovative perspectives on the role of information professionals in this emerging field.

Finally, the resources were organized into thematic categories to facilitate comparative analysis. These categories included: theoretical foundations of big data, applications in scientific research, competencies and roles of information professionals, and ethical and technical challenges. This structuring allowed for a deeper and more systematic understanding of the topic, contributing to the development of a relevant and well-grounded theoretical framework.

Definition and history of Big Data

The term “big data” first appeared in 1997 in an article by NASA researchers Michael Cox and David Ellsworth (1997). In this article, the authors highlighted the difficulties of storing, processing, and managing the increasingly large volumes of data required for their research with the technologies of the time.

Despite this initial emphasis on volume as an identifying element of big data, with the development and expansion of the Web and other data-generating technologies and services, its characterization became more complex. Doug Laney (2001), a computer engineer at the Gartner consulting firm, added two more “Vs” to the definition of the concept: velocity and variety. Velocity refers to the real-time creation of data (e.g., social network content), while variety points to the fact that data can come from different sources, in different formats, and at different levels of structuring. Subsequently, other authors (Bansal et al. 2020) have suggested incorporating other “Vs” into the characterization of big data, such as veracity, value, and variability. Lupton (2015) even proposed replacing all these Vs with 13 Ps that consider the sociocultural dimensions of big data: portentous, perverse, personal, productive, partial, practices, predictive, political, provocative, privacy, polyvalent, polymorphous, and playful.

However, as Kitchin and McArdle (2016) have shown, there is no characteristic profile that fits all big data, but rather different “types” of big data. There are datasets that are very large but not created in real time, and conversely, there are datasets created in real time that are not very large. After examining the nature of 26 datasets described in the literature as big data, the authors conclude that the two characteristics that truly differentiate big data from small data are velocity and exhaustiveness, understood as the ability to collect data from the entire system rather than just a sample.

Finally, it is worth noting that while initially the data collected was mainly related to the physical world of applied sciences, over time—and especially with the advent of social networks—the scope of data has expanded into social science research. Big data on human behavior has become a new source of knowledge for many researchers, unimaginable just a few decades ago. The use of this new resource raises sensitive questions related to data privacy and ethics in handling personal information.

Impact of Big Data on Scientific Research

The big data explosion has obviously had a huge impact on the field of academic research. In the last 15 years, every discipline has, to differing degrees, seen quantitative and qualitative changes to the possibilities of research work and the methodologies used to conduct it (Abdalla et al. 2024; Tosi et al. 2024).

Initially oriented towards computer science, finance, healthcare, and marketing, the scope of big data applications has expanded over the years to include disciplines as diverse as biology (ElSayed et al. 2021), astronomy (Klapp et al. 2024), education (Fischer et al. 2020), and the social sciences (Chen 2018).

As early as 2007, Jim Gray, a Microsoft researcher and a pioneer in database software, pointed out the significant impact that computing was having on the practice of science, a transformation that Gray called a “fourth paradigm” (Lynch 2009), represented by the integration of the three previous paradigms (theory, experimentation, and large-scale computation) in a context of data-intensive science. Bell (2009) traces the characteristics of the fourth paradigm back to Johannes Kepler’s exploitation of Tycho Brahe’s astronomical observations catalog for the discovery of the laws of planetary motion.

Based on this premise, the main opportunities that big data offers scientific research include:

- Acceleration of research: the availability of vast amounts of data and the ability to analyze it quickly and efficiently makes it possible to obtain results that previously required months or years of data collection and processing. Examples include the discovery of new drugs (Latha Bhaskaran et al. 2022), optimization of agricultural productivity (Misra et al. 2022), and support for disease diagnosis (Leopold et al. 2020).
- Emergence of new research fields: big data has enabled many scientists to tackle challenges that were previously insurmountable. In the field of the social sciences there are many examples due to the availability of extensive data on human behavior. Whether through data generated by social networks (Angskun et al. 2022) or by sensors (e.g., detecting people’s movement in a public space) (Jang et al. 2024), scientists can now formulate hypotheses and predictions that were previously unimaginable.
- Interdisciplinary collaboration: the complexity of capturing, organizing, and exploiting big data necessarily requires collaboration with computer scientists and statisticians at a minimum. Beyond this “forced” collaborative work, the capacity for exploiting big data allows many researchers to tackle projects drawing on related data sources from different fields. In public health, the spread of infectious diseases is often analyzed using big data from social networks, electronic medical records, and mobile services (Huang et al. 2020).

Similarly, in the book value chain —encompassing publishers, booksellers, and librarians— big data helps improve efficiency in the supply chain, increase the discoverability of books, and better understanding readers' needs to offer them the best content and services (Attanasio 2022).

However, these opportunities are not without challenges for researchers. As will be discussed below, some of these challenges can be addressed by information and documentation professionals:

- Data discovery: finding out whether the necessary data for a research project exists and where it is located is not always an easy task. Although there are an increasing number of repositories operated by governments (Data.gov, datos.gob.es, data.gov.in, Dades obertes, etc.), international organizations (World Bank Open Data, UNData, etc.), and institutions (Re3Data), and even thematic repositories (Dryad) and dataset search engines (Google's Dataset Search), many scientists are unaware of them or find them difficult to use.
- Data characteristics: beyond locating the data, rigorous research requires knowing the basic characteristics of the data: who created it, when it was created and for what purpose, what tools and methods were used to collect it, what organization published it, what fields make up the dataset, what formats it was published in, what access rights apply to it, and any other information necessary for its correct use and interpretation. Gathering data (especially from platforms such as social networks) raises important ethical and privacy challenges that must be addressed from the very beginning of the data lifecycle. In this regard, the data scientist has a key responsibility to ensure that data has been collected transparently, respectfully, and with full awareness of the rights of the individuals involved. It is not only about protecting data once it has been stored but about adopting an ethical stance from its origin. Without comprehensive information, scientists are at risk of basing their research on irrelevant data. Hence the importance of describing datasets with internationally consolidated metadata standards such as Dublin Core or DataCite Schema.
- Data quality: once it has been located and its relevance has been verified, scientists need to check the quality of the data. It is estimated that 80% of a data scientist's work is spent on cleaning processes (Subirats Maté et al. 2019; Ciavotta et al. 2022), the most common tasks being removing duplicates, correcting errors in data types and forms, normalizing controlled values, and ensuring the integrity of missing data. Additionally, if data is being used from different sources, data cleaning becomes even more important, as the goal of achieving a unique and coherent dataset can be hindered by the different structures and forms of each source. Beyond technical considerations, it is essential to critically assess data representativeness, especially in the field of social sciences and official data. Large volumes of data do not, by themselves, ensure a comprehensive or equitable view of reality, as they often originate from biased sources, exclude certain social groups, or reflect structural inequalities. This lack of representativeness can distort analyses and lead to misleading conclusions if not critically addressed. Therefore, Information and Documentation professionals must be aware of these limitations and promote data collection and analysis practices that embrace diversity and inclusion.
- New knowledge and skills: for scientists, the main challenge of working with big data is not its volume but the need to acquire knowledge and skills specific to data science that were often not required before. While it is true that developing advanced knowledge and

skills in statistics (probability, inference, multivariate analysis, etc.), mathematics (linear algebra, calculus and numerical analysis, graph theory, etc.), and computing (programming languages such as R or Python, databases, data visualization, etc.) can be a titanic task for a single person, it is equally true that a researcher intending to work with big data needs at least a basic grounding in these areas to form part of a multidisciplinary team and work side by side with statisticians, mathematicians, and computer scientists.

- New infrastructures: the three basic Vs of big data (volume, velocity, and variety) require specific computing architecture. To efficiently store, process, and analyze large volumes of data (in some cases generated in real time), systems with scalable or distributed storage capacity (computer clusters) and intensive processing (multiple CPU cores and large RAM capacity) are needed. High-speed networks are also advisable for the rapid transfer of data between system components. Obviously, this type of architecture comes at a cost that only large research centers can afford, and while cloud computing services can reduce the size of this investment, it is still significant (Abdalla 2022; Abdalla et al. 2024).
- Data security and privacy management: while security refers to all the tools and technologies necessary to monitor any type of attack, theft attempt, or other type of security breach (Das et al. 2021), privacy specifically relates to the handling of personal data, which is especially sensitive in research related to fields such as healthcare, finance, or social networks. It is important for researchers to be familiar with tools and techniques such as firewalls, antivirus software, anonymization, and encryption, so that they can work together with cybersecurity specialists to develop a comprehensive security plan for the project and understand the legal framework in which it operates (the General Data Protection Regulation in the case of the European Union). Cases such as the Cambridge Analytica scandal (Lapowsky 2018) highlight the potential impact of improper or malicious handling of personal data.

The Role of Information and Documentation Professionals

In the context described above, Information and Documentation professionals can play a crucial role. Their experience and skills in searching for, organizing, storing, analyzing, and managing information mean that they could make a valuable contribution to the management of the big data generated in research (Gbaje 2024).

According to Ohaji, Chawner, and Yoong (2019), data librarians play an essential role in the improvement of data accessibility and preservation in academic libraries and research institutions. Their professional profile enables them to offer services such as secure data storage, advice on open data policies, and training in data management. Additionally, their user-oriented focus places them in a privileged position to adapt to the objectives, needs, and limitations of any project. Garoufallou and Gaitanou (2021, 421) and Morriello (2020) have highlighted the importance of Big Data in the evolution of library services, the challenges that still need to be addressed, and the need for new skills for librarians.

Table 1 presents an outline of these skills, highlighting capabilities that are essential for responsible data management, such as effective communication, mastery of technological tools, and familiarity with intellectual property regulations and licenses.

Skills	Description
Interpersonal and behavioral characteristics	Formal communication skills, e.g., writing technical documents and case studies. This also includes adaptability to new experiences and always keeping updated on new developments.
Contextual knowledge of the institutional environment	Knowledge of funding policies of agencies that support scientific research. This includes an understanding of ethical procedures in the scientific process, disciplinary research methods, scientific communication, intellectual property, access methods, copyright, Creative Commons licenses, and royalties.
Data usage	An understanding of types of data (quantitative, qualitative), metadata schemas (Dublin Core, DataCite Schema), unique identifiers (DOI), and digital data preservation.
Knowledge of data manipulation technologies	Technical knowledge that includes an understanding of programming languages (Python, SQL, Java), database design and structuring, APIs for data retrieval, user-centered design, natural language processing, the Internet of things, and big data.
Conceptual skills	An understanding of data management principles, including data collection, organization, storage, preservation, and access. Knowledge of ontologies and controlled vocabularies to model and represent knowledge in a structured way. The ability to communicate complex information through effective data visualizations. Basic knowledge of artificial intelligence techniques such as machine learning and natural language processing to extract value from data.
Management	The ability to plan, organize, and execute complex data management projects. Effective communication skills to interact with users from different disciplines and explain technical concepts clearly and concisely. The ability to work in multidisciplinary teams and collaborate with other information professionals, researchers, and developers. Leadership skills to guide and motivate teams in data management projects.

Table 1. Skills of Information and Documentation professionals for handling big data in the context of scientific research.
Source: Based on Semeler and Pinto (2020). Available at: <https://www.scielo.br/j/tinf/a/LNskQZxLgCVPRvbFqN-9vZz/?lang=en#>

The skills of Information and Documentation professionals mentioned above qualify them to take on a range of tasks aimed at handling data in the context of scientific research, where it is possible to establish various methodologies and techniques for studying data, taking into account its nature and type. According to Morriello (2020, 3), in some countries, associations such as CILIP (United Kingdom) have identified data science as a key area for librarians, highlighting competencies in data management and curation, machine learning, and data analysis. Similarly, entities like the Australian National Data Service (ANDS) have outlined a precise profile for the data librarian, emphasizing their role in managing and preserving research data. Under this premise, Figure 1 presents the tasks that an information and documentation professional can perform in the field of data management within the context of scientific research.

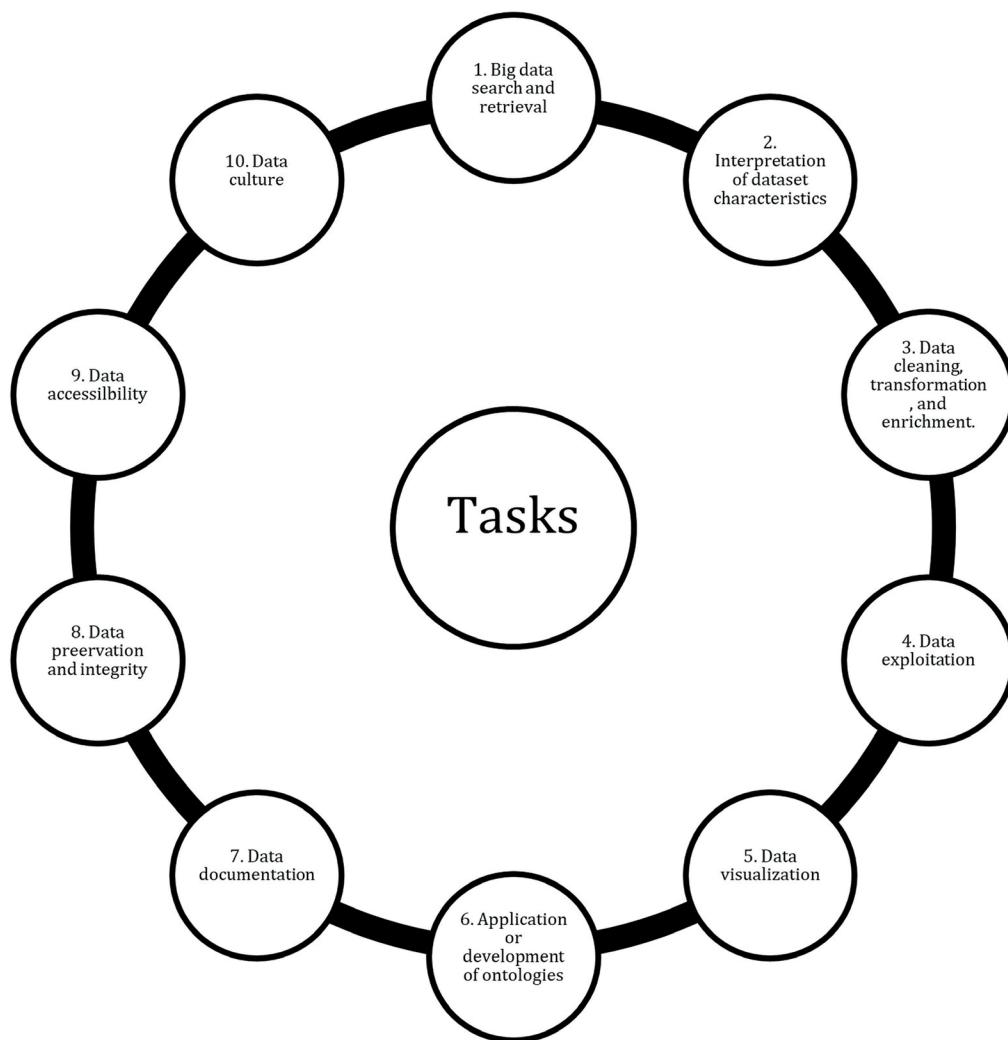


Figure1. Tasks of the Information and Documentation professional for data management in the context of scientific research. Source: own elaboration.

These tasks include the following:

1. Big data search and retrieval. The current abundance of data and the variety of available sources (from open repositories and institutional databases to social networks and restricted access sources) have created a significant challenge for researchers. The diversity of formats, data quality, access levels, and platforms is making the retrieval process increasingly complex. This context demands information and documentation professionals with advanced skills in search techniques, the use of Boolean operators, data mining, and the selection of specialized tools to optimize retrieval when dealing with masses of information (Chapman et al. 2020; Morriello 2020).
2. Interpretation of dataset characteristics. With the development of advanced metadata schemas such as DataCite Schema or Data Catalog Vocabulary (DCAT), datasets can include a very detailed level of description (e.g., authorship, creation context, applied methodologies,

formats, and limitations), which is key to their interpretation and proper use. However, this wealth of information is often too vast for the researcher to interpret every element and use it in their analysis. Here, the information and documentation professional plays a crucial role in breaking down and contextualizing these metadata, facilitating interpretation and ensuring that the data is accessible and understandable to the research community. According to Sanjuán (2019, 7), analyzing data always involves reflecting on it, regardless of the research project's methodological approach.

3. Data cleaning, transformation, and enrichment. Raw data generally suffers from problems such as duplicates, empty fields, typographical errors, and inconsistent values that hinder its analysis and exploitation. Using tools like OpenRefine, information and documentation professionals can perform exhaustive cleaning processes to correct and homogenize the data. These tools also allow the integration of additional data through external API queries (such as Wikidata) to enrich the information with specific details (e.g., geolocation, thematic categories, or unique identifiers), maximizing its value and usefulness in multidisciplinary research (Ciavotta et al. 2022; Ridzuan and Zainon 2019).
4. Data exploitation. Exploiting big data requires not only a basic understanding of the data itself but also advanced skills in statistics and programming to extract patterns, perform predictive analyses, and apply methods such as data mining or machine learning. An information and documentation professional with knowledge in these fields can act as a key advisor or collaborator, helping researchers derive meaningful conclusions from complex datasets and discover relationships that might otherwise go unnoticed. In recent years, numerous methods, techniques, and technologies have been developed to help organizations make better data-based decisions and extract value from the data they collect. These are used for capturing, processing, storing, managing, and analyzing data in order to exploit it more effectively and obtain greater benefits from it (Curto 2016).
5. Data visualization. The interpretation of large volumes of data is greatly facilitated by visualization. Tools like Tableau and Power BI allow the creation of graphs, maps, and interactive dashboards that translate complex data into clear and understandable visual representations. With a mastery of these tools, information and documentation professionals can design intuitive visualizations that not only effectively communicate findings but also allow researchers to interact with the data and explore various perspectives. According to Tufte (2007), the visual representation of data helps avoid misinterpretations. In this respect, graphs must be faithful to the data and avoid visual manipulations that distort reality.
6. Application or development of ontologies. Ontologies offer a valuable means of structuring and relating data in a way that is readable by both humans and machines. An ontology can be used to classify data and connect different datasets, allowing researchers to identify underlying relationships, to make inferences, and to automate analysis processes. This approach is especially useful in the field of artificial intelligence and natural language processing, where data must be semantically understood to make complex discoveries. An ontology models the knowledge of a particular domain, which could be useful in another application; effective sharing of ontologies is therefore very important for data reuse (Alfaifi 2022).

7. Data documentation. In an environment where massive data is intended to be reused, proper documentation is essential. Creating accurate descriptions based on standardized metadata schemas such as DataCite Schema, Data Catalog Vocabulary (DCAT), or CEDAR, or on more generic ones like Dublin Core or Schema.org (Guha, Brickley, and Macbeth 2016), ensures that data is accessible, understandable, and reusable in the future (Ávila Barrientos 2020). This correct and precise identification of datasets facilitates their citation in scientific texts, in addition to supporting a variety of data management activities and the development of reproducible, transparent, ethical, and high-impact research practices. Moreover, to foster a culture of data integration, scientists need to be convinced that preparing their data for online publication is an effort worth making (Brase 2010).
8. Data preservation and integrity. Data curators should also implement strategies that ensure the longevity and sustained integrity of data, in order to guarantee its validity and relevance for future research (Morriello 2020). This work is essential to ensure that published data has the same level of value, reliability, and accessibility as a conventional academic publication. According to Condon, Simpson, and Emanuel (2022), research data integrity is crucial for scientific rigor, and “although [the] inability to reproduce or replicate a study’s findings is not always a research data integrity issue, it is an important factor in successful attempts.” Furthermore, data reuse depends on the integrity of research data and the reliability of the original dataset.
9. Data accessibility. In the context of data repository management, information and documentation professionals have to design and develop digital systems to store and share research data securely and efficiently. They need to ensure that repositories function properly and are updated regularly, and to implement access and use policies that will guarantee data protection while promoting its reuse. The FAIR principles (findability, accessibility, interoperability, reusability) highlight the need to make data and other digital assets more accessible. Such accessibility is a key fundamental principle for bringing data into a domain with clearly defined access rules. There are three main components to the application of this principle: access protocol, access authorization, and metadata longevity (Wise et al. 2019).
10. Data culture. There is a need to promote a data culture by raising awareness in the scientific community about the importance of data management and open science. This can be done by providing researchers with training on the development of good data management practices and fostering collaboration between researchers and information and documentation professionals to optimize the data lifecycle. Data policies, mandates, and initiatives at the international level reveal the value of research data, as well as the importance of being able to share and reuse it. Ultimately, the promotion of a data culture gives meaning to any institutional action to coordinate the work carried out by researchers (Melero 2018; Morriello 2020).

In addition, Attanasio (2022) points out that the emergence of big data has led to a transformation in the way metadata is generated, shifting from a model where a human validates relationships between entities to a system where machines create connections by tracking digital events. This automation presents an opportunity to improve information management but also raises challenges in terms of data control and ethics.

Therefore, the data librarian emerges as a key figure in research and teaching, playing a crucial role in the curation and management of research data. However, the impact of big data and artificial intelligence has redefined how data and metadata are generated and used.

Conclusions

Information and documentation professionals play a pivotal role in the efficient management of massive data within the landscape of contemporary scientific research. Their expertise in organizing, structuring, and curating extensive datasets ensures not only their quality and accessibility but also enhances the reproducibility and integrity of research outcomes. In an environment increasingly driven by complex data analysis, these professionals contribute significantly to the success of multi-, inter-, and transdisciplinary projects by enabling seamless information flow and facilitating informed decision-making.

Beyond supporting data search and retrieval, their responsibilities extend to ensuring long-term data preservation and fostering its ethical and responsible reuse. By implementing rigorous metadata standards, data governance frameworks, and interoperability practices, they help mitigate risks associated with data loss, misinterpretation, and security breaches. Additionally, their role in promoting open science principles strengthens collaborative networks and democratizes access to knowledge, fostering innovation across diverse fields of inquiry.

To remain at the forefront of these evolving demands, information and documentation professionals must continuously develop competencies in data management, digital preservation, and emerging technologies. Mastery of artificial intelligence, machine learning, and data visualization techniques will allow them to provide strategic insights and add substantial value to research in the era of big data. By embracing these challenges, they position themselves as indispensable partners in the advancement of knowledge and scientific discovery.

The article provides a broad overview of the opportunities, challenges, skills, and tasks faced by Information and Documentation Professionals (IDPs) in the context of big data and scientific research. While the proposed lists are valuable and comprehensive, there is a need to further explore their practical application and to establish priorities within each category.

One relevant line of development involves using the skills list as a basis for empirical studies aimed at identifying training gaps in specific contexts. This could support the design of professional development programs more closely aligned with actual workplace demands.

Similarly, the task list provides a useful framework for investigating, through surveys or field studies, the real roles played by IDPs in concrete settings such as research centers, academic libraries, or data management units. This approach would allow for an assessment of whether the roles defined in theory match those practiced in everyday professional contexts.

Finally, it is recommended that future versions of the analysis be structured according to strategic priorities, distinguishing between critical, desirable, and emerging tasks. This would not only facilitate its use by institutional managers but also help IDPs align their professional development with key competencies for the contemporary digital environment.

References

- Abdalla, Hemn Barzan. 2022. "A Brief Survey on Big Data: Technologies, Terminologies and Data-Intensive Applications." *Journal of Big Data* 9 (1): 107. <https://doi.org/10.1186/s40537-022-00659-3>.
- Abdalla, Hemn Barzan, Ardalan Hussein Awlla, Yulia Kumar, and Maryam Cheraghy. "Big Data: Past, Present, and Future Insights." In *CTCNet '24: Proceedings of the 2024 Asia Pacific Conference on Computing Technologies, Communications and Networking*, 1–15. Chengdu, China, September 2024. <https://doi.org/10.1145/3685767.3685777>.
- Alfaifi, Yasser. 2022. "Ontology Development Methodology: A Systematic Review and Case Study." In *2nd International Conference on Computing and Information Technology (ICCIT)*, January 25–27, 446-50. <https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9711664>.
- Almeida, Maria B., Renato R. Souza, and Renato M. A. Baracho. 2015. "Looking for the Identity of Information Science in the Age of Big Data, Computing Clouds and Social Networks." In *Proceedings of the 4th International Symposium of Information Science (ISI)*, Croatia. https://www.researchgate.net/publication/273448682_Looking_for_the_identity_of_Information_Science_in_the_age_of_big_data_computing_clouds_and_social_networks.
- Angskun, Jiratchaya, Siriporn Tipprasert, and Thara Angskun. 2022. "Big Data Analytics on Social Networks for Real-Time Depression Detection." *Journal of Big Data* 9 (1). <https://doi.org/10.1186/s40537-022-00622-2>.
- Attanasio, Pietro. 2022. "New Challenges in Metadata Management between Publishers and Libraries." *JLIS.it* 13 (1): 1–22. <https://doi.org/10.4403/jlis.it-12777>.
- Ávila Barrientos, Eder. 2020. "Análisis de los Principios para la Descripción de Datos de Investigación mediante DataCite Metadata Schema." *Anales de Documentación* 23 (1): 1–14. <https://doi.org/10.6018/analesdoc.400341>.
- Bansal, Mohit, Inderveer Chana, and Siobhán Clarke. 2020. "A Survey on IoT Big Data: Current Status, 13 V's Challenges, and Future Directions." *ACM Computing Surveys* 53 (6): 1–59. <https://doi.org/10.1145/3419634>.
- Bell, Gordon. 2009. "Foreword." In *The Fourth Paradigm: Data-Intensive Scientific Discovery*, edited by Tony Hey, Stewart Tansley, and Kristin Tolle, XI–XV. Redmond, Washington: Microsoft Research.
- Bhaskaran, Krithika Latha, Richard Sakyi Osei, Evans Kotei, Eric Yaw Agbezuge, Carlos Ankora and Ernest Domanaanmwii Ganaa. 2022. "A Survey on Big Data in Pharmacology, Toxicology and Pharmaceutics." *Big Data and Cognitive Computing* 6 (4): 161. <https://doi.org/10.3390/bdcc6040161>.
- Bräse, Jan. 2010. "DataCite: A Global Registration Agency for Research Data." In *RatSWD Working Paper*, 149. Berlin: Rat für Sozial- und Wirtschaftsdaten (RatSWD). <https://www.econstor.eu/bitstream/10419/43624/1/638240131.pdf>.
- Chapman, Adriane, Elena Simperl, Laura Koesten, George Konstantinidis, Luis-Daniel Ibáñez, Emilia Kacprzak, and Paul Groth. 2020. "Dataset Search: A Survey." *VLDB Journal* 29 (1): 251–72. <https://doi.org/10.1007/s00778-019-00564-x>.

Chen, Shu-Heng, ed. 2018. *Big Data in Computational Social Science and Humanities*. Cham: Springer.

Ciavotta, Michele, Vincenzo Cutrona, Flavio De Paoli, Nikolay Nikolov, Matteo Palmonari, and Dumitru Roman. 2022. "Supporting Semantic Data Enrichment at Scale." In *Technologies and Applications for Big Data Value*, 19–39. Cham: Springer. <https://doi.org/10.1007/978-3-030-78307-5>.

Condon, Patrick, Jennifer Simpson, and Michael Emanuel. 2022. "Research Data Integrity: A Cornerstone of Rigorous and Reproducible Research." *IASSIST Quarterly* 46 (3). <https://doi.org/10.29173/iq1033>.

Cox, Michael, and David Ellsworth. 1997. "Managing Big Data for Scientific Visualization." *ACM SIGGRAPH* 97 (1): 21–38. https://www.researchgate.net/profile/David-Ellsworth-2/publication/238704525_Managing_big_data_for_scientific_visualization/links/54ad79d20cf2213c-5fe4081a/Managing-big-data-for-scientific-visualization.pdf.

Curto, Joan. 2016. *Fundamentos de Big Data: Habilitando la Explotación de Datos Complejos*. UOC. https://openaccess.uoc.edu/bitstream/10609/136326/2/Fundamentos%20y%20usos%20del%20big%20data_Mo%C2%BFdulo2_Fundamentos%20del%20big%20data%20Habilitando%20la%20explotaci%C2%BFn%20de%20datos%20complejos.pdf.

Das, Pradeep Kumar, Harihar K. Tripathy, and Siti Aisyah M. Yusof. 2021. *Privacy and Security Issues in Big Data: An Analytical View on Business Intelligence*. Cham: Springer.

ElSayed, Ibrahim A., Kamal ElDahshan, Hoda Hefny, and Eman K. ElSayed. 2021. "Big Data and Its Future in Computational Biology: A Literature Review." *Journal of Computer Science* 17 (12): 1222–8. <https://doi.org/10.3844/jcssp.2021.1222.1228>.

Fischer, Christian, Zachary A. Pardos, Ryan S. Baker, Joseph Jay Williams, Padhraic Smyth, Renzhe Yu, Stefan Slater, Rachel Baker, and Mark Warschauer. 2020. "Mining Big Data in Education: Affordances and Challenges." *Review of Research in Education* 44 (1): 130–60. <https://doi.org/10.3102/0091732X20903304>.

Garoufallou, Emmanouel, and Panorea Gaitanou. 2021. "Big Data: Opportunities and Challenges in Libraries, a Systematic Literature Review." *College and Research Libraries* 82 (3): 409–28. <https://crl.acrl.org/index.php/crl/article/view/24918/32750>.

Gbaje, Ezra. 2024. "Big Data Analysis: An Emerging Paradigm in Information Management." *Nigerian Library Association Kogi State Chapter 2021 Conference*, Kogi State University Ayingba, Kogi State, September 6. https://www.researchgate.net/publication/384838489_Big_Data_Analysis_An_Emerging_Paradigm_in_Information_Management.

Guha, Ramanathan V., Dan Brickley, and Steve Macbeth. 2016. "Schema.org: Evolution of Structured Data on the Web." *Communications of the ACM* 59 (2): 44–51. <https://doi.org/10.1145/2844544>.

Huang, Haibo, Zhaohui Peng, Haibo Wu, and Qiong Xie. 2020. "A Big Data Analysis on the Five Dimensions of Emergency Management Information in the Early Stage of COVID-19 in China." *Journal of Chinese Governance* 5 (2): 213–33. <https://doi.org/10.1080/23812346.2020.1744923>.

Jang, Seongmin, Juhyun Bae, and Younghoon Kim. 2024. "Street-Level Urban Heat Island Mitigation: Assessing the Cooling Effect of Green Infrastructure Using Urban IoT Sensor Big Data." *Sustainable Cities and Society* 100. <https://doi.org/10.1016/j.scs.2023.105007>.

Kitchin, Rob, and Gavin McArdle. 2016. "What Makes Big Data, Big Data? Exploring the Ontological Characteristics of 26 Datasets." *Big Data and Society* 3 (1). <https://doi.org/10.1177/2053951716631130>.

Klapp, Jaime, Celia R. Fierro-Santillán, and Leonardo G. Sigalotti. 2024. "El Estudio de las Estrellas en la Era del Big Data." *CIENCIA ergo-sum* 31. <https://doi.org/10.30878/ces.v31n0a36>.

Laney, Doug. 2001. "3D Data Management: Controlling Data Volume, Velocity and Variety." *META Group Research Note*, 6.

Lapowsky, Issie. 2018. "Facebook Exposed 87 Million Users to Cambridge Analytica." *Wired*, April 4. <https://www.wired.com/story/facebook-exposed-87-million-users-to-cambridge-analytica/>.

Leopold, Jane A., Bradley A. Maron, and Joseph Loscalzo. 2020. "The Application of Big Data to Cardiovascular Disease: Paths to Precision Medicine." *Journal of Clinical Investigation* 130 (1): 29–38. <https://doi.org/10.1172/JCI129203>.

Liao, Hsin-Ta, Zhijie Wang, and Yang Liu. 2020. "Exploring the Cross-Disciplinary Collaboration: A Scientometric Analysis of Social Science Research Related to Artificial Intelligence and Big Data Application." *IOP Conference Series: Materials Science and Engineering* 806 (1). <https://doi.org/10.1088/1757-899X/806/1/012019>.

Lupton, Deborah. 2015. "The Thirteen Ps of Big Data." *The Sociological Life*, May 13. <https://simplysociology.wordpress.com/2015/05/11/the-thirteen-ps-of-big-data/>.

Lynch, Clifford. 2009. "Jim Gray's Fourth Paradigm and the Construction of the Scientific Record." In *The Fourth Paradigm: Data-Intensive Scientific Discovery*, edited by Tony Hey, Stewart Tansley, and Kristin Tolle, 177–83. Redmond, Washington: Microsoft Research.

Melero, Remedios. 2018. *Recomendaciones para la Gestión de los Datos de Investigación*. Maredata. <https://digital.csic.es/handle/10261/173801>.

Misra, N. N., Yash Dixit, Ahmad Al-Mallahi, Manreet Singh Bhullar, Rohit Upadhyay, and Alex Martynenko. 2022. "IoT, Big Data, and Artificial Intelligence in Agriculture and Food Industry." *IEEE Internet of Things Journal* 9 (9): 6305–6324. <https://doi.org/10.1109/JIOT.2020.2998584>.

Morriello, Rosanna. 2020. "Birth and Development of Data Librarianship." *JLIS.it* 11 (3): 1–15. <https://doi.org/10.4403/jlis.it-12653>.

Ohaji, Ifeanyi K., Brenda Chawner, and Pak Yoong. 2019. "The Role of a Data Librarian in Academic and Research Libraries." *Information Research* 24 (4). <https://informationr.net/ir/24-4/paper844.html>.

Ridzuan, Fakhitah, Wan Mohd Nazmee, and Wan Zainon. 2019. "A Review on Data Cleaning Methods for Big Data." *Procedia Computer Science* 161: 731–38. <https://doi.org/10.1016/j.procs.2019.11.177>.

Rossi, Renata, and Kenji Hirama. 2015. "Characterizing Big Data Management." *Issues in Informing Science and Information Technology* 12: 165–80. <http://iisit.org/Vol12/IISITv12p165-180Rossi1921.pdf>.

Sanjuán Núñez, Luis. 2019. *El Análisis de Datos en Investigación Cualitativa*. UOC. https://open-access.uoc.edu/bitstream/10609/147145/2/MetodosDeInvestigacionCualitativaEnElAmbitoLaboral_Modulo5_ElAnalisisDeDatosEnInvestigacioCualitativa.pdf.

Semeler, Alexandre R., and Ana L. Pinto. 2020. "Data Librarianship as a Field Study." *Transinformação* 32. <https://doi.org/10.1590/2318-0889202032e200034>.

Subirats Maté, Lluís, Daniel O. Pérez Trenard, and María Calvo González. 2019. *Introducción a la Limpieza y Análisis*. UOC. <https://openaccess.uoc.edu/bitstream/10609/148647/1/IntroduccionALaLimpiezaYAnalisisDeLosDatos.pdf>

Tosi, Davide, Rezon Kokaj, and Marco Roccetti. 2024. "15 Years of Big Data: A Systematic Literature Review." *Journal of Big Data* 11 (1). <https://doi.org/10.1186/s40537-024-00914-9>.

Tufte, Edward R. 2007. *The Visual Display of Quantitative Information*. Cheshire: Graphic Press.

Wise, John, Alexandra Grebe de Barron, Andrea Splendiani, Beeta Balali-Mood, Drashti Vasant, Eric Little, Gaspare Mellino, Ian Harrow, Ian Smith, Jan Taubert, Kees van Bochove, Martin Rommeker, Peter Walgemoed, Rafael C. Jimenez, Rainer Winnenburg, Tom Plasterer, Vibhor Gupta, Victoria Hedley. 2019. "Implementation and Relevance of FAIR Data Principles in Biopharmaceutical R&D." *Drug Discovery Today* 24 (4): 933–8. <https://doi.org/10.1016/j.drudis.2019.01.008>.