



Archives, linked data and the digital humanities: increasing access to digitised and born-digital archives via the semantic web

Ashleigh Hawkins¹ 

Accepted: 10 December 2021 / Published online: 27 December 2021
© The Author(s) 2021

Abstract

Mass digitisation and the exponential growth of born-digital archives over the past two decades have resulted in an enormous volume of archives and archival data being available digitally. This has produced a valuable but under-utilised source of large-scale digital data ripe for interrogation by scholars and practitioners in the Digital Humanities. However, current digitisation approaches fall short of the requirements of digital humanists for structured, integrated, interoperable, and interrogable data. Linked Data provides a viable means of producing such data, creating machine-readable archival data suited to analysis using digital humanities research methods. While a growing body of archival scholarship and praxis has explored Linked Data, its potential to open up digitised and born-digital archives to the Digital Humanities is under-examined. This article approaches Archival Linked Data from the perspective of the Digital Humanities, extrapolating from both archival and digital humanities Linked Data scholarship to identify the benefits to digital humanists of the production and provision of access to Archival Linked Data. It will consider some of the current barriers preventing digital humanists from being able to experience the benefits of Archival Linked Data evidenced, and to fully utilise archives which have been made available digitally. The article argues for increased collaboration between the two disciplines, challenges individuals and institutions to engage with Linked Data, and suggests the incorporation of AI and low-barrier tools such as Wikidata into the Linked Data production workflow in order to scale up the production of Archival Linked Data as a means of increasing access to and utilisation of digitised and born-digital archives.

Keywords Digital archives · Linked data · Digital humanities · Cultural heritage · Artificial Intelligence

✉ Ashleigh Hawkins
a.h.hawkins@liverpool.ac.uk

¹ Liverpool University Centre for Archive Studies, Liverpool, UK

Introduction

The mass digitisation of archives and exponential growth in born-digital archives over the past two decades, together with the publication of large volumes of aggregated cultural heritage metadata over the past decade, have resulted in an enormous volume of archives and associated (meta)data, being available digitally. This has produced a valuable but under-utilised source of large-scale digital data ripe for interrogation by scholars and practitioners in the interdisciplinary field of the Digital Humanities, opening up digitised archives to a range of novel digital research methods. Using computational methods, digital humanists interrogate large digital datasets to address an ever-expanding range of humanities questions. This article has adopted a broad interpretation of the term “the Digital Humanities” as being at the intersection of computational technologies and humanities scholarship (Schreibman et al. 2004), and uses the term “digital humanist” to refer to any person engaged in scholarship or practice within the field of Digital Humanities and/or using computational methods for humanities research. The Digital Humanities is driven by digital data; however, it is not enough that archival data is available digitally, it needs to be integrated, interoperable, and interrogable (Bikakis 2021; Koho et al. 2021; Zeng 2019). Current approaches to digitisation—the process of producing a digital surrogate of an analogue object—usually result in the production of unstructured data, for example in the form of a digital image, sound or moving image file, consequently presenting a number of barriers to digital humanists.

In order to increase the utilisation of digitised and born-digital archives by digital humanists, it is essential that their contents and associated metadata are made available in a machine-readable format. Such “datafication” “...renders a diverse range of information as machine-readable, quantifiable data for the purpose of aggregation and analysis” (Southerton 2020). Digitisation is often a primary stage in datafication, yet as a process it remains distinct from datafication, which emerged with the growth of data analytics and big data over the past decade (Mayer-Schönberger and Cukier 2013). It is beyond the scope of this article to provide a detailed examination of datafication. However, it must be noted that the process of datafication has been critiqued as losing much qualitative detail of the source material (Southerton 2020), as perpetuating and subjecting new harms and discrimination upon already vulnerable communities (Noble 2018; Sutherland 2019), and as having “benefited governments, medical institutions, and corporations at the expense of citizens’ and consumers’ liberty and privacy” (Schüll 2020, pp. 457), among other criticisms. On the other hand, Blanke and Prescott assert that datafication has particular significance for the humanities as it “suggests that for big data work unstructured data is not enough” (2016, pp. 193). Michael Moss et al. suggest the key consequence of the move towards datafication for archives is that they have been transformed from a collection of administrative records into “a collection of data to be mined” (Moss et al. 2018, pp. 120).

When engaged with critically, the datafication of digitised archives provides many new opportunities for digital humanists to make use of the vast quantity

of archives digitised over the past two decades. However, making such heterogeneous and distributed digital data available in accordance with the FAIR Guiding Principles—that data be published as Findable, Accessible, Interoperable and Re-usable data (Wilkinson et al. 2016)—is one of the major challenges facing the cultural heritage domain. Linked Data provides a viable means of archival datafication capable of implementing the FAIR Guiding Principles, creating machine-readable, interoperable, extensible Archival Linked Data suited to interrogation and analysis using digital humanities research methods. Using Linked Data, archival data (catalogue data, metadata, data extracted from the contents of born-digital and digitised archives) can be embedded into the web, enriching and further contextualising archival data, and making it easier to discover, access, and utilise.

Soon after the introduction of Linked Data, early case studies began to explore its potential application in the archives sector (Clough et al. 2011; Ruddock 2011). In the intervening decade, a substantial body of associated research has continued to develop examining the application of Linked Data to archives. Numerous published case studies are available, with Karen F. Gracy and Jinfang Niu providing some scholarly assessment. The majority of available literature principally focuses on the technical challenges posed by the production of Linked Data and the development of solutions to them. Critical engagement and assessment of Archival Linked Data is currently absent, and there has been little consideration of the benefits that Archival Linked Data offers to digital humanities research and scholarship, or of how it can increase utilisation of digitised or born-digital archives. In digital humanities scholarship, in contrast, there is growing evidence of how Linked Data specifically supports research. Over the past decade an increased adoption of Linked Data is evident in digital humanities scholarship, where it has been presented as a means of increasing the interoperability of digital cultural heritage data (Jordanous et al. 2012; Miyakita et al. 2018). Case studies provide evidence to suggest that Linked Data benefits the Digital Humanities in that it provides a viable means of publishing the digital data which fuels the Digital Humanities in the form of integrated, harmonised, and interoperable large-scale datasets which can be reused in multiple ways. However, notwithstanding that evidence, analysis of Linked Data and how it benefits the Digital Humanities remains limited.

Despite an evident growth of interest, Linked Data remains under-examined in both archival and digital humanities scholarship. Scholarly examination, has placed a heavy emphasis on the technical aspects—the “how” of Linked Data, but has given little consideration of the “why”, i.e. the benefits of Linked Data. This article begins to address this gap by examining how Archival Linked Data is beneficial for those engaged in digital humanities research and scholarship. In considering some of the barriers currently preventing digital humanists from being able to utilise digitised and born-digital archives, and suggesting the production of Archival Linked Data as part of interdisciplinary collaboration as a way forward, this article is a first step towards developing an understanding of the potential intersections between archival and digital humanities Linked Data scholarship and praxis.

It does this in four ways: first, after a very brief introduction to Linked Data, the article provides a snapshot of the Linked Data landscape in both archival and

digital humanities scholarship. Secondly, extrapolating from archival scholarly, case-study, and practice-based literature, it details the benefits of Archival Linked Data which directly support increasing access to digitised archives for the purposes of digital humanities scholarship and practice. Thirdly, it examines some of the barriers currently preventing digital humanists from being able to take full advantage of the abundant potential of Archival Linked Data, and of utilising current digitised or datafied archives. Drawing these findings together, the author considers future directions of Archival Linked Data, including its role in processing and increasing access to born-digital archives, the incorporation of Artificial Intelligence (AI) and low-barrier tools to scale up the production of Archival Linked Data, and increased collaboration between archival and digital humanities scholars and practitioners. This is a theoretical account; further research is needed to validate, through real-life application, the benefits of Linked Data for the purposes of digital humanities scholarship. In particular, examination of how the technological and structural barriers highlighted here are actually affecting the critical work of, for example, a national archive institution, and identification of the changes and developments necessary to ensure the progress of digital scholarship dependent on Linked Data, would be especially valuable contributions to current scholarship and practice.

Linked Data

First conceived in 2006 as the building blocks of the Semantic Web, Linked Data is most simply described as a “set of best practices for publishing and connecting structured data on the Web” (Bizer et al. 2009, pp. 1). Whereas the Web is based upon documents (e.g. web pages) being linked to other documents, the Semantic Web is structured in the form of semantically defined data linked to other data, known as Linked Data. Linked Data itself is not a specific technology or tool, rather it is a set of principles for publishing structured data which underpin the creation of interoperable machine-readable data on the Semantic Web. The Linked Data Principles state:

1. Use URIs [Unique Resource Identifiers] as names for things
2. Use HTTP URIs so that people can look up those names
3. When someone looks up a URI, provide useful information, using the standards (RDF [Resource Description Framework], SPARQL [SPARQL Protocol and RDF Query Language])
4. Include links to other URIs so that they [people and intelligent Semantic Web agents] can discover more things (Berners-Lee 2006)

Linked Data is structured in the form of triples as subject-predicate-object, whereby the predicate semantically defines the relationship between the two entities. By using unique identifiers, standardised ontologies and vocabularies, large, interconnected networks of these entities and their relationships are created, which can be represented in the form of a knowledge graph.

The Linked Data principles were followed up in 2010 by the 5 Star Linked Open Data Scheme which incorporated the Open Knowledge Foundation’s definition of Open Data as that which “can be freely used, modified and shared by anyone for any purpose” (Open Knowledge Foundation (ND) 2021). The 5 Star Scheme defines a sliding scale of the fundamental technical and accessibility requirements of Linked Data as it moves towards becoming Linked Open Data: data which is available on the web (1*), as machine-readable structured data (2*), in a non-proprietary format (3*), using standards of the World Wide Web Consortium (RDF and SPARQL) (4*), and linked to other Linked Open Data (5*) (Berners-Lee 2010). This article has adopted the term Linked Data to refer collectively to Linked Data that is published under an open licence (i.e. Linked Open Data) and that which is not. A distinction has only been maintained when discussing individual case studies where the language of the author(s) has been used, and for resources which include Linked Open Data in their title. For an in depth introduction to Linked Data see Gracy 2015 (Fig. 1).

Archives and Linked Data

The year 2011, five years after Tim Berners-Lee’s first articulation of the Linked Data Principles, is a significant turning point in the emergence of interest in Linked Data within the archives sector. Although the Library of Congress published authority files as RDF triples in 2008 (Summers et al. 2008), it was in 2011 that reports on the potential uses and benefits of Linked Data for Libraries, Archives and Museums first emerged (Isaac et al. 2011; Keller et al. 2011). The same year saw the publication of the first complete datasets of bibliographic Linked Data—the British Library’s British National Bibliography, and Oslo Public Library Linked Data

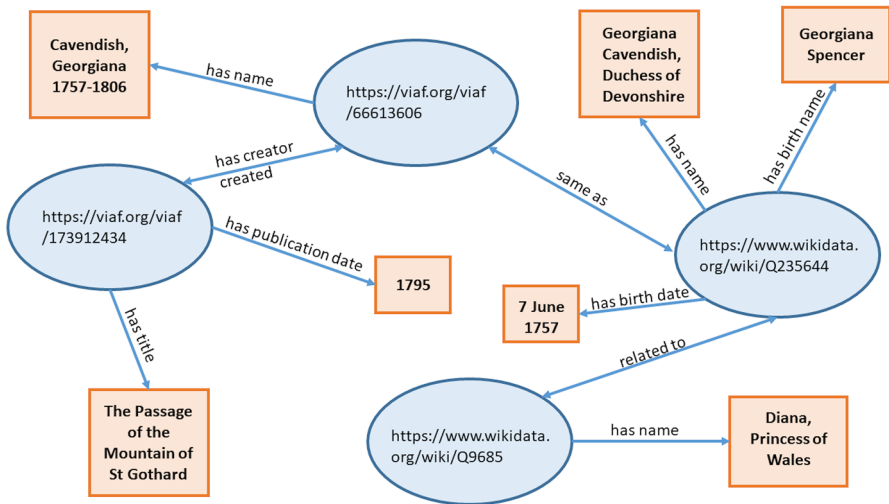


Fig. 1 A simple knowledge graph

(Godby et al. 2015; Rekkavik 2014)—and early articles reporting on archives-specific Linked Data projects, such as research at the UK National Archives on geolocation referencing (2009–2011) and the LOCAH (Linked Open COPAC and Archives Hub) project (2010–2011) (Clough et al. 2011; Ruddock 2011). One of the first archival instances of a Linked Data dataset derived from the LOCAH Project, which published a subset of archival metadata taken from the Archives Hub, an archival metadata aggregator, together with bibliographic data from Copac (Consortium of Online Public Access Catalogues) (Ruddock 2011). Through identifying a number of challenges that the project team was facing, including issues of sustainability, maintenance of provenance, and lack of standardisation of archival data, the LOCAH project raised awareness of the specific barriers facing archive services in the adoption of Linked Data. On the other hand, it also made clear the numerous benefits which Linked Data could bring—including integration with other data sources, the facilitation of serendipitous search, and the creation of flexible entry points to archival data, thereby increasing the exposure of collections.

Interest has continued to grow over the following decade, as evidenced by the plethora of published case studies now available. These case studies generally fall into one of three categories. Firstly, those which develop Archival Linked Data infrastructure, such as gazetteers (Clough et al. 2011), mark-up schemas (Dobreski et al. 2019; Gartner 2015), tools and workflows (Pattueli et al. 2013; Stevenson 2012a) and ontologies (Corn and Patrick 2019; Llanes-Padrón and Pastor-Sánchez 2017; Robledano-Arillo et al. 2020). Most notable in this category is the ontology of the International Council on Archives' latest standard for archival description, Records in Contexts, which has been presented as the Linked Data ontology, RiC-O (Pitti and Stocking 2016). The second category, which is most common, is those which enhance discrete pre-existing collections of archival metadata and publish them as Linked Data, including for the Historical Archive of the European Commission (Damova 2020), the Zeri Photo Archive (Daquino et al. 2017), and the Getulio Vargas Foundation's Archives (Rademaker et al. 2015).

In the third category are new digital archival resources made available via Linked Data web services and user interfaces. An early example is the Online Digital Archive of Architectural Practice in Post-War Queensland, a digital archive providing a comprehensive online resource of oral history interviews—supplemented with items from related archive and library collections (Hunter et al. 2012). The case study identified a number of benefits of adopting Linked Data in the development of digital archives, providing evidence that benefits can be experienced equally by archive services, their key stakeholders, and users. The Linked Jazz project created a Linked Open Data dataset of over two thousand jazz musicians extracted from oral history interviews (Pattueli et al. 2013, 2015, 2017). The dataset is available via a network visualisation tool which enables analysis of the professional and social relationships between individual musicians. Linked Jazz presented a convincing argument to support further investment in Archival Linked Data: “As the amount of digital cultural heritage data continues to grow at an exponential rate, there is a call for new strategies and applications to enhance their discovery, interpretation, and use. The application of LOD [Linked Open Data] technology to cultural heritage

content holds enormous potential to answer this call” (Pattueli et al. 2013, pp. 6) (see Fig. 2).

Archival scholars examining Linked Data include Gracy’s analysis of the opportunities and challenges of implementing Linked Data (2015), and experiments using Linked Data to increase access to library metadata for music resources and moving image archive metadata (Gracy et al. 2013, 2018). By developing crosswalks between datasets, Gracy has also provided tools for the implementation of Linked Data, and made a number of recommendations to overcome the challenges associated with the processes involved (2015). Niu’s scholarly overview of early trends in archival Linked Data practice revealed that the majority of projects up to 2015 had primarily converted existing descriptions. Furthermore, she found that the implementation of Linked Data significantly changes archival description, and argued that many of the methods of accessing Linked Data are beyond the capabilities of, what she terms, “generic users” of archives (Niu 2016).

The Digital Humanities and Linked Data

As is the case in archival scholarship, the trend in digital humanities Linked Data scholarship has been for individual case studies. Writing in 2012, the authors of a case study using Semantic Web techniques to enrich TEI (Text Encoding Initiative) encoded manuscripts, argued that “the benefits of semantic web techniques are currently under-explored in Digital Humanities research”, positioning Linked Data as a solution to data being made available but not interoperable (Jordanous et al. 2012, pp. 44). Since then, case studies have demonstrated how various digital humanities research methods can be combined with Linked Data to support research in a range of disciplines, including Linguistics (Chiarcos et al. 2018; Cimiano et al.

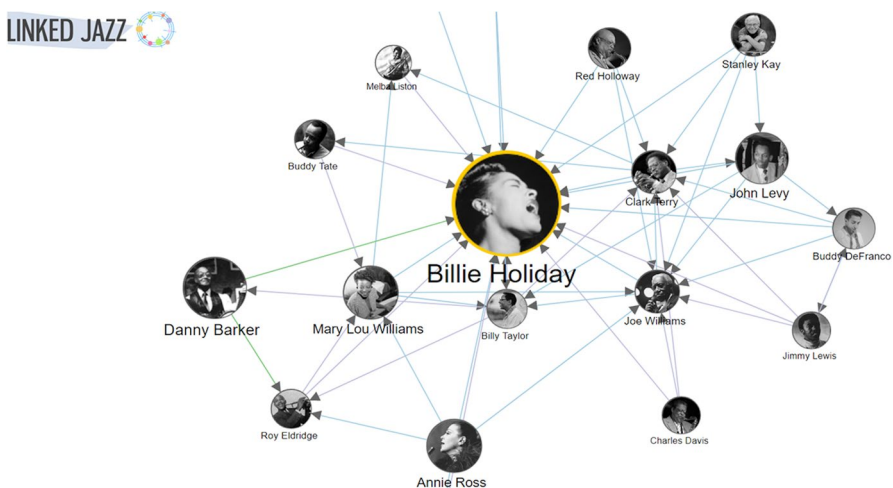


Fig. 2 Linked Jazz project by the Semantic Lab at Pratt Institute—network visualisation tool (CC BY)

2020), Literary Studies (Egloff and Picca 2020; García et al. 2016), Music (Eyharabide et al. 2019), and Digital History (Balado et al. 2015; Blanke and Riechert 2020; Bruneau et al. 2021; Horne 2020; Hyvönen et al. 2016; Miyakita et al. 2018; Romein et al. 2020; Tamper et al. 2018).

Of particular relevance, are the case studies combining digital humanities methods with digital cultural heritage data; as a recent overview has argued: “Linked Data and Semantic Web technologies are becoming increasingly important in creating, publishing, and analysing Cultural Heritage data in Digital Humanities” (Bikakis et al. 2021, pp. 166). Earlier case studies focused on creating a discrete Linked Data dataset and examining the links between data within the dataset, or with a specific external dataset. The Sharing Ancient Wisdoms project, for example, created a digital edition of Greek and Arabic collections of ancient wise sayings using RDF. The project demonstrated the value of incorporating Linked Data into the process of producing online scholarly editions marked up using the TEI XML format, and identified advantages of that approach for digital humanists wishing to create similar digital editions, or to analyse the data in resulting datasets (Jordanous et al. 2012). The Dutch Ships and Sailors project developed and evaluated a method for identifying links between ships in two datasets, a dataset of Dutch maritime events, the “Northern Muster Rolls database”, and the historical newspaper archive of the Dutch National Library (Balado et al. 2015). The project combined machine learning techniques with manual assessment to identify links between the two datasets, proving not only the possibility of retrieving a considerable amount of relevant links between the two datasets, but also that the links provided new opportunities for analysis of the source datasets.

However, as more and more interoperable datasets have become available, an increasing number of studies have focused on the development of tools, user interfaces, and web services to allow for the access of Linked Data from multiple sources without the need for advanced technical expertise (Baierer et al. 2017; Egloff and Picca 2020; Hoekstra et al. 2016; Miyakita et al. 2018; Szytler et al. 2014; Xia et al. 2018). A key example is DIVE, a Linked Data digital cultural heritage collection browser which uses historical events and narratives as the context for searching, browsing, and providing access to objects from heterogeneous collections (de Boer et al. 2015). A particularly well-known example is WarSampo—Finnish Second World War on the Semantic Web, an initiative which integrated nineteen distributed datasets relating to Finland in the Second World War as Linked Open Data, and created the WarSampo Data for Digital Humanities Linked Open Data service to enable research and the creation of applications related to war history (Hyvönen et al. 2016; Koho et al. 2021). Additionally, the initiative has developed the WarSampo Portal to provide access to those without Linked Data skills, enabling search and browsing based on originally six, and subsequently increased to nine, different “perspectives”, including event, place, person, casualty, and army unit (Hyvönen et al. 2016; Koho et al. 2021). WarSampo has transformed and harmonised previously isolated datasets to form a unified and extensible knowledge graph which enables queries that were hitherto impossible, including presenting biographies for individual soldiers and histories of military units (Koho et al. 2016, 2021) (see Fig. 3).

The screenshot shows the WarSampo website interface. At the top, there is a navigation bar with 'Perspectives', 'View Help', 'Settings', and 'Suomeksi'. Below this, the main content area is titled 'Persons' and 'Einar August Vihma E. Wichmann'. On the left, there is a search box with the text 'Search for known persons from the past Finnish wars by writing their name in the text input below and/or selecting a person from the list below. Information regarding the person and recommended links will appear on the right. If you cannot find the person you are looking for, and know in which military unit they have served, you can take a look at the unit's timeline.' Below the search box, a list of results is shown, with 'Vihma E. Wichmann, Einar August (Major General)' selected. The main content area for Einar August Vihma E. Wichmann includes tabs for 'Information', 'Timeline', and 'Photographs'. The 'Information' tab is active, showing a brief biography: 'Einar August Vihma (vuoteen 1936 Wichmann; 19. syyskuuta 1893 Kuopio – 5. elokuuta 1944 Ihtantala) oli suomalainen kenraalimajuri ja Mannerheim-ristin ritari. Talin-Ihtantalan taistelussa kuoliut Vihma oli viimeinen jatkosodassa kaatunut jääkäri ja tietyvästi itsenäisen Suomen ainoa sodassa rintamalla kaatunut kenraali. Muita sodassa kuolleita kenraaleja ovat Kaarlo Viljanen ja Aarne Snelman, jotka ylennettiin kenraaleiksi vasta haavoittumisen jälkeen. Vihma osallistui myös 1912 Tukholman kesäolympialaisten joukkuevalmistukseen varuinenä saaden olympiahopeaa. (Wikipedia)'. Below the biography is a 'Personal Details' table with the following information:

Family name	Vihma (ent. Wichmann) ^[R] Vihma E. Wichmann
Given names	Einar August ^[R]
Born	19.09.1893 ^[R]
Municipality of birth	Kuopio ^[R]
Municipality of domicile	Helsinki ^[R]
Municipality of residence	Helsinki ^[R]
Citizenship	Finnish ^[R]
Nationality	Finnish ^[R]
Mother tongue	Finnish ^[R]
Marital status	Married ^[R]

On the right side of the page, there is a 'Vihma, Einar' section with a birth and death date: 'Born on 1893 in Kuopio. Died on 1944 in Ihtantala.' Below this, there is a 'divisionankomentaja' section with a detailed biography: 'Einar Vihma muistetaan paljolti viimeisenä jatkosodassa kaatuneena jääkäriin; samalla hän oli itsenäisen Suomen ainoa sodassa rintamalla kaatunut kenraali. Suomalaisessa sota- ja sotahistoriassa hänellä on kuitenkin paikkansa myös jääkärikojien etappi- ja joukko-osastokomentajana sekä ennen muuta yhtenä talvi- ja jatkosodan merkittävistä rintamakomentajista. Source: Semantinen Kansallisbiografia \ SKS:n Kansallisbiografia'.

Fig. 3 WarSampo Finnish Second World War on the Semantic Web—faceted search interface (CC BY)

Indeed, Finland has become a key centre in developing and investing in the Linked Open Data infrastructure required to provide access to digital cultural heritage data for the purposes of the Digital Humanities. To date this has included the development of an ontology service, Linked Data publishing platform, and various cultural heritage Semantic Web portals (Hyvönen 2020a). These portals, including BiographySampo, BookSampo, CultureSampo, NameSampo, TravelSampo and the above-mentioned multi-award winning WarSampo, have been accessed by millions of end-users, including specialist digital humanities scholars, genealogists and the general public (Hyvönen 2020a; Hyvönen et al. 2019; Koho et al. 2016, 2021). The Sampo portals, together with a number of other case studies, clearly demonstrate that investment in Linked Data tools and services for the Digital Humanities benefits users in general, and aids the development of cultural heritage applications more widely (Blanke and Riechert 2020; de Boer et al. 2015; Hyvönen 2020a; b; Hyvönen et al. 2016; Koho et al. 2021; Oldman et al. 2016; Tamper et al. 2018).

The benefits of Archival Linked Data for Digital Humanities scholarship

Whereas digital humanities scholarship demonstrates an interest in, and awareness of, the value of Archival Linked Data for the Digital Humanities, there has been little similar consideration of its value, and its benefits, within archival scholarship—this despite the fact that, at least in one instance, inspiration for the project came directly from the Digital Humanities to the archive sector. In a study developing a Linked Data informed ontology for the description of Spanish Civil War photographic archives, the authors acknowledged that the context of their research was partially informed by “the transformations in the organization of documentary collections spurred by the digital humanities” (Robledano-Arillo et al. 2019, pp. 67). However,

the aim of the study itself was directed to supporting historians using traditional research methods in accessing photographs for iconographic research. Beyond this, there are some general suggestions that in processing large datasets, revealing relationships across heterogeneous sources, and enabling access to unstructured data, Archival Linked Data offers opportunities for digital humanities research (Bones 2019; Pattuelli et al. 2017). However, what exactly these opportunities and benefits are remains under-examined.

Notwithstanding the lack of *explicit* investigation of the benefits of Archival Linked Data for the Digital Humanities, it is evident that many of the benefits of Linked Data to archive services, their collections, and users which have been articulated from within the archival sector are especially advantageous to the Digital Humanities. Archival Linked Data meets the needs of digital humanists in that it provides interoperable, reusable, and integrated data. However, in making archival data machine-readable, connecting disparate and cross-disciplinary datasets, and revealing previously unknown relationships across collections, the benefits of Archival Linked Data for digital humanities scholarship extend much further, and deserve to be much more clearly articulated across the disciplines.

Within archival scholarship, the literature contains multiple examples of Archival Linked Data's relevance and value to the Digital Humanities community. Improvements in knowledge discovery, information search and retrieval greatly benefits digital humanists due to improved effectiveness, efficiency, and precision (Llanes-Padrón and Pastor-Sánchez 2017; McKenna et al. 2018; Rademaker et al. 2015; Robledano-Arillo et al. 2019). More complex queries can be accommodated through the use of SPARQL. With a single SPARQL query, searches can be made across multiple collections, and navigation across cultural heritage and non-cultural heritage sources of data is made possible (Gracy 2015; Gracy et al. 2013; McKenna et al. 2018). Linked Data makes it easier to reuse, align and enrich archival data (Pitti et al. 2016), and integrate it with data derived from other sources (Clough et al. 2011; Gracy 2015; Hunter et al. 2012; Niu 2016; Pattuelli et al. 2017). In this way, the data cleaning and harmonisation processes which occupy a significant part of any digital humanities research that brings together data from multiple sources are made redundant. However, it must be noted that this is in large part due to these processes having already been undertaken in the production of Linked Data, and which constitute one of the most time-consuming aspects of the process (Davis 2019). Linked Data pushes archival data closer to the individual user, allowing them to more efficiently access data for their specific applications (Debruyne et al. 2016; Rademaker et al. 2015), and to navigate seamlessly between datasets (Baierer et al. 2017; Clough et al. 2011; McKenna et al. 2018). In accommodating multiple search methods, Linked Data enables different search methods to be made available to meet the needs of digital humanists with a range of technical abilities and research interests. These include SPARQL endpoints, entity, semantic concept or keyword search, browsing and exploration, and serendipitous search.

As a machine-readable format, Linked Data is capable of supporting automatic reasoning and analysis of semantic data, querying large volumes of data, and offering new methods for discovery, engagement, interpretation, and use. Web-based user interfaces often incorporate tools which facilitate new methods of engagement and analysis, and “allow the use of archival data in ways that a few years ago were

unimaginable or prohibitively difficult to do for both social and technological reasons” (Pitti et al. 2016, pp. 176). Adopting Linked Data creates a digital research environment ideal for digital humanists, opening up archival data to natively digital methods, and supporting dynamic research methods (Gartner 2015; McKenna et al. 2018). Furthermore, Linked Data extends the depth and breadth of archival analysis possible as it can be interrogated through, for example, graphical interfaces, data or text mining, data clustering, information visualisation, network analysis, natural language processing (NLP), and named entity recognition (NER), many of which digital methods are commonly employed in the Digital Humanities.

The process of integrating archival datasets and making them interoperable increases the quality of the data made available for the Digital Humanities as it is necessary to adhere to cross-domain standards and links are created across sources of data. Improvements in quality are manifold: archival data is richer, more expressive and granular, and semantically enriched (Browell 2016; Gartner 2015; Gracy et al. 2013; Niu 2016; Pattuelli et al. 2013; Pitti et al. 2016). The integration and harmonisation of heterogeneous sources of data facilitates automatic reasoning, providing the ability to generate new knowledge and infer additional implicit facts from those explicitly documented (Damova 2020; Tillman 2016). Archival data is also enriched through incorporating knowledge derived from both users and experts. Users can add their own knowledge to archival description through the incorporation of user-generated description into Linked Data datasets with activities such as semantic tagging and other user annotations (Gracy 2015; Niu 2016). Collection descriptions are enhanced through the provision of additional archival data, such as further descriptive, contextual, and authority data which might be internal administrative data not previously publicly available, or data drawn from external sources (Gartner 2015; Gracy 2015, 2018). Data is, therefore, better contextualised by it not being presented in isolation, enabling digital humanists to more easily gain a fuller understanding of archival data.

Current barriers to Digital Humanities scholarship

The challenges of producing and publishing Archival Linked Data are relatively well documented, and include technological issues, a prevalence of unstructured and not easily disambiguated data, a lack of financial and skilled human resources, low-levels of awareness of Linked Data within the profession, and an absence of an Archival Linked Data infrastructure, including tools, standards and best practice (Gracy 2018; Hyvönen 2012; McKenna et al. 2018; Smith-Yoshimura 2018). Challenges for which both archival and digital humanities scholars and practitioners share an impetus in finding a solution include: balancing the provision of access to open data with the maintenance of Intellectual Property Rights (Gartner 2015; McKenna et al. 2018), preventing the decontextualisation and loss of nuance of archives (Gartner 2015; Gracy 2018), and providing access without complicating the user search process (Niu 2016). As the growing body of case studies has demonstrated, many challenges can be overcome with the maturation of Archival Linked Data practices. Moreover, recent large-scale projects have begun to deliver some of the fundamental

infrastructure required for engagement with the production of Archival Linked Data en masse. These projects include the development of an international archives sector Linked Data ontology, RiC-O, already referred to above, and, as part of the UK Towards a National Collection project, investigation of the infrastructure required for the creation and maintenance of Persistent Identifiers (PIDs) in the UK heritage sector, and exploration of Linked Open Data, knowledge graphs, and AI as a means of making connections between online representations of heritage objects and information about them.¹

It is clear that Archival Linked Data provides valuable advantages for digital humanists wishing to interrogate digitised archives. However, Linked Data is far from a panacea to the challenges of accessing the contents of digitised archives, and their associated metadata, as data. To date, Archival Linked Data is a minor player on the Linked Data scene. Furthermore, even though more is being contributed year on year, it is still a statistically insignificant proportion of existing digital archival data globally. In addition to this limited availability of archival data as Linked Data, a key hindrance to the success of Archival Linked Data for the purposes of digital humanities scholarship is the current lack of relevant archival and non-archival Linked Data datasets with which archival data can be interconnected and augmented.

The Linked Open Data Cloud provides a visualisation of all datasets currently available as Linked Open Data (i.e. as RDF, linked via URIs to other datasets, and with open access to the entire dataset) and the links between them. Composed (as at May 2021), of 1,301 datasets, it is dominated by datasets from the life sciences, linguistics and governmental data. At the centre of the Cloud are DBpedia² and Wikidata,³ two crowd-sourced Linked Data datasets created as part of Wikimedia Foundation projects, and GeoNames, a multilingual geographical database.⁴ Archival and bibliographic datasets are mostly included in the Publications Subdomain where key datasets include VIAF (the Virtual Internet Authority File),⁵ which links to twenty-eight other datasets, the Library of Congress Subject Headings which is linked to twenty-seven datasets,⁶ and the European digital cultural heritage metadata aggregator Europeana, which links to six.⁷ A number of Archival Linked Data datasets are included, such as the National Digital Data Archive of Hungary,⁸ the Bibliothèque nationale de France,⁹ and the 20th Century Press Archives.¹⁰ The former two datasets link to both DBpedia and VIAF, among others, while the latter links

¹ Towards a National Collection, <https://www.nationalcollection.org.uk/>.

² DBpedia, <https://www.dbpedia.org/>.

³ Wikidata, https://www.wikidata.org/wiki/Wikidata:Main_Page.

⁴ GeoNames, <https://www.geonames.org/>.

⁵ VIAF, <http://viaf.org/>.

⁶ Library of Congress Subject Headings, <https://id.loc.gov/authorities/subjects.html>.

⁷ Europeana Pro, <https://pro.europeana.eu/page/linked-open-data>.

⁸ Linked Open Data Cloud: National Digital Archive of Hungary, <https://lod-cloud.net/dataset/sztaki-lod>.

⁹ data.bnf.fr, <https://data.bnf.fr/en/about>.

¹⁰ 20th Century Press Archives, <https://lod-cloud.net/dataset/zbw-pressemappe20>.

to DBpedia and GeoNames. Such Archival Linked Data datasets confirm the possibility of building meaningful connections between Archival Linked Data and other Linked Data datasets, thus providing impetus for the addition of further Archival Linked Data to the Linked Open Data Cloud.

However, available Linked Data datasets present a number of challenges when identifying relevant datasets which correspond to, and can enrich, digitised archives, partially as a result of the fragmented transnational and corporate infrastructures affecting the scope of datasets and their recorded information. One dataset which illustrates many of these issues is, GeoNames, a commonly used digital gazetteer in both archival and digital humanities Linked Data activity requiring historic geographic data. GeoNames integrates global geographic data and makes it available as Linked Data, including place names in multiple languages, latitude and longitude coordinates, and current population statistics. Although this is undoubtedly useful information for, say, the mapping of historic data, it provides little contextual information of relevance to historic places and spaces. Similarly, trying to integrate person-centred historic data is also not straightforward. VIAF is limited to information about published authors; Wikimedia Foundation projects are governed by a notability test which requires people to be considered “worthy of notice”, “significant”, or “remarkable” to warrant inclusion (Wikipedia 2021). Such requirements led to pages created to document enslaved people owned by Martha and George Washington at Dogue Run Farm not being published by Wikipedia due to an absence of relevant published source material (Herbert and Parilla 2021). It is clear, therefore, for the majority of individuals documented in digitised historical archives—who may be peasants or workers, prisoners or wives and mothers—little or no related content is available as Linked Data. As well as resulting in links only being possible between a limited number of individuals, the narrow scope of these key Linked Data datasets risks prolonging the privileging of a minority of “notable” individuals captured in the now digitised historic record, further marginalising historically excluded communities.

Fortunately, relevant digital data is available. The collections of digitised archives made accessible as a result of mass digitisation projects provide almost limitless potential for the Digital Humanities. Often driven by the needs of family historians for sources of names, census records, vital records (birth, death, marriage), title deeds, wills, and many other record types have been digitised and made publicly available in the form of digital images, and/or data, created through indexing and the use of Optical Character Recognition (OCR). The inadequacy of these approaches to digitisation for the purposes of the Digital Humanities, however, are captured by Sonia Ranade, who, reflecting on the digitisation work of the UK National Archives, states.

In early digitisation projects, our emphasis (and that of most other archival institutions) was facilitation of remote access, creating digitised assets with the assumption that these will be consumed in much the same way that an on-site researcher reads a paper file. In effect, our traditional digitisation activities deliver a “picture” of the page, alongside some limited indexing. We have found that digitised resources created in this way do not readily lend them-

selves to computational analysis and we are beginning to realise that our established approach to digitisation does little to facilitate innovative use of the records.

(Ranade 2016, pp. 3264).

Furthermore, increased commercialisation of the digitisation of archives has left much of this vast resource of potential data trapped behind the paywalls of academic publishers and commercial genealogy companies. Such commercialisation has not only dictated the types of record prioritised for digitisation, it has also influenced how the resulting data is formatted. For example, trying to meet the demand of genealogists has led to the publication of data in forms which enable search for individuals, rather than the examination of total populations (Morris 2017).

Through datafication, the potential of this largely untapped resource could be released, opening up digitised archives to digital methods of interaction, interrogation, and analysis. A number of large-scale projects have demonstrated the value of datafying digitised archives to form large databases combing digitised archives from multiple sources, with notable examples from the UK including University College London's Legacies of British Slave-ownership project, which created the online Encyclopaedia of British Slave-ownership,¹¹ and People of Medieval Scotland, which has created a database of named individuals drawn from over 8,600 documents covering the period 1093–1371.¹² One key example is the Digital Panopticon, a cross-institutional and international digital humanities collaboration, which has created a database relating to people sentenced at the London Old Bailey between 1780 and 1875.¹³ The project brought together genealogical, biometric, and criminal justice datasets held by multiple organisations based in the UK and Australia. The project also created new digital archival datasets by digitising and datafying further historic records, manually combining and harmonising digitised archives from multiple sources. The problem with the creation of such databases is that, neither open nor connected to external data, such databases are commonly structured as relational databases and provide only mediated access via a bespoke search tool, often developed specifically for the project. Frequently, such databases do not provide access to the underlying data, limiting its exploitation to within the parameters defined by the original project. Furthermore, when made available for download, considerable data-cleaning and harmonisation is required to enable integration with other datasets. The addition of such a wealth of data to the Linked Open Data Cloud would be a major boon far beyond digital humanists.

From this perspective, then, there are a number of obstacles to overcome, and there currently seems to be little viable alternative to the commercial model of mass digitisation. Furthermore, Thomas Padilla points to the role that cultural heritage organisations play in “enclosing” their own data through commercial partnerships, questioning “Is it worth boosting the cost of admission to an enclosed garden that

¹¹ Centre for the Study of the Legacies of British Slavery, <https://www.ucl.ac.uk/lbs/>.

¹² People of Medieval Scotland, <https://www.poms.ac.uk/>.

¹³ Digital Panopticon, <https://www.digitalpanopticon.org/>.

weakens the library community and inhibits emerging forms of research”)? (2018, pp. 297). One way forward might be individual projects negotiating free access with commercial providers to subsets of their data, as was the case of the Digital Panopticon which negotiated free access to aspects of the key genealogical data held by their commercial providers, Findmypast and Ancestry. Indeed, the impact of access restrictions to digitised archives on Archival Linked Data activity can be seen from the Irish Record Linkage 1864–1913 project which used digitised archives and data provided by the General Register Office “under strict terms and conditions” which limited access only to the project team (Debruyne et al. 2016, pp. 160). While such a compromise is far from the ideal solution—in the case of the Digital Panopticon, subscriptions are still required for full access to the transcriptions and digitised images from which the data has been extracted—the partnership model of digitisation funding, from commercial partners, internal funding, and grant funding, has become a relatively common one.

Future directions of Archival Linked Data?

Engagement with Linked Data has come a long way over the past ten years in both archival and digital humanities scholarship and practice. The previous review has indicated how, to date, Archival Linked Data case studies have focused primarily on increasing access to archival metadata or, more rarely, digitised archives, and a number of barriers remain to making a significant proportion of these materials available as Linked Data. Digital humanities case studies have focused on increasing the accessibility of Archival Linked Data and the development of tools for access and analysis. However, those currently available are yet to accommodate the wide range of digital humanities research topics and methods required (Hyvönen 2020b). Resolving the challenges of making digitised archives both available and accessible as Linked Data are complex. It not only requires technological developments and mass investment in the production of Linked Data by multiple sectors, but also a change in attitude towards data being open and freely available (within legal limits), in addition to an infrastructure that enables this. The FAIR Guiding Principles can aid the cultural heritage domain move in this direction and are highly compatible with Linked Data. Guidelines for the application of these principles to the digital objects and metadata held by libraries, archives and museums have been proposed, including the use of PIDs, standardised metadata formats, and external-links (Koster and Woutersen-Windhouwer 2018).

Significantly, digitised archives are only one category of the digital archives currently available. The exponential growth of the creation of born-digital records since the 1990s is set to vastly increase the volume of born-digital archives. The huge volume of digital archives requires archivists to no longer consider archives as texts to be read but as collections of data, for which new methods and tools are required to undertake a reformed approach to appraisal (Moss et al. 2018). While progress has been made in the preservation and provision of access to born-digital archives it has not been rapid enough: “born-digital archives are endangered archives, and we urgently need to preserve these collections, make them

available and produce new knowledge... It is astonishing that email and born-digital archives are still treated as a new thing that few archivists really understand” (Jaillant 2019, pp. 300). Moreover, traditional search methods based on free-form natural language search are ineffective in exploring large born-digital collections (Winters and Prescott 2019). The potential of Linked Data as a means of making born-digital archives available as collections of data, accessible to multiple users and research methods, warrants further investigation. Lucy McKenna, in a paper presented at the AURA Workshop, suggested the benefits of Linked Data for born-digital archives included improved knowledge discovery, seamless navigation, and increased awareness of available resources in digital archives (2020). Whereas it is likely that many of the practices for the creation and provision of access to Archival Linked Data originating from digitised archives will be transferable to the context of born-digital archives, there will inevitably be some areas of difference and additional challenges. Linked *Open* Data is preferable for the purposes of digital humanists, yet the format is not always appropriate for archives where restrictions relating to sensitivity, confidentiality, data protection, and Intellectual Property Rights are not uncommon. Such restrictions are especially prevalent in born-digital archives and have resulted in a vast amount of born-digital archives currently being preserved but not made available for access. Tools to combat these challenges are beginning to emerge, particularly for use with email collections (Jaillant 2019; Ries and Palkó 2019; Schneider et al. 2019). For example, ePADD, an open-source software developed by Stanford Libraries’ Department of Special Collections and University Archives, incorporates machine learning and natural language processing to screen email collections for sensitive and legally restricted information, prepare emails for preservation, and make them discoverable and accessible to users (Schneider et al. 2019). As a recent editorial argued, while it is essential that the preservation of born-digital records is accelerated, “we also need to push for access to these archives through lobbying for open data respectful of privacy” (Jaillant 2019, pp. 301). One approach to this has been demonstrated by the 20th Century Press Archives which donated a large Linked Data dataset to Wikidata in 2019. By publishing only the metadata created by the 20th Century Press Archives under an open licence (CC0), the organisation found a way of contributing data regarding archives with access restrictions to the Linked Open Data cloud—providing a controlled level of access to their contents—while continuing to uphold the intellectual property rights to the original material (Neubert 2019). Linked Data, both that which is open and “closed” (i.e. made available with access and/or use restrictions), offers an as yet under-explored potential to be part of the solution to the challenge of providing some level of access to born-digital archives while maintaining certain restrictions.

Despite the scale of these challenges, the expertise of archivists and other cultural heritage professionals has an important role to play in overcoming the issues currently preventing the widespread adoption of Linked Data in the archives sector (McKenna et al. 2018). Even while the necessary infrastructure remains absent for the production, provision of access to, maintenance, and preservation of Archival Linked Data, archivists can make a significant contribution at an institutional or

individual level. By increasing their own knowledge of Linked Data, and improving institutional awareness and preparedness for Archival Linked Data, archive services will be well placed to embrace Archival Linked Data when it becomes a more feasible prospect. Investing in data cleaning and standardisation, and exploring low-barrier means of contributing archival data as Linked Data, such as Wikidata, are achievable first steps for many archive services which can have a considerable impact on increasing access to high-quality archival data and ensuring its future interoperability.

The adoption of Linked Data in the archives sector is still a relatively new phenomenon and is open to expansion and development in multiple directions. In the library sector, a number of reports have been published in recent years concerning and advocating the use of Wikipedia and Wikidata (Association of Research Libraries 2019; International Federation of Library Associations and Institutions 2017; Program for Cooperative Cataloguing 2018). The library sector's increasing engagement with Wikimedia Foundation projects may prophesize a future development of Archival Linked Data infrastructure. In providing a low-barrier entry to the Linked Data environment capable of scaling-up Linked Data adoption, Wikidata may prove a key tool in overcoming some of the outstanding infrastructural challenges. Cultural heritage institutions which have drawn in authority data from Wikidata to their datasets and reciprocally contributed their data to Wikimedia Foundation projects include the National Galleries of Scotland (Gould 2018), Museum of Modern Art (MOMA) (Romeo 2016), the Library of Congress (Thornton 2017), and WorldCat (Proffitt 2020). The advantages of making cultural heritage data freely accessible via Wikidata are clearly demonstrated by the Smithsonian's experience: the amount of views of an identical image is on average 1000 times higher on Wikipedia than on the Smithsonian's own web pages (Kapsalis 2019).

The increasing attention given to AI in both the Digital Humanities and archival scholarship (as elsewhere) suggests a further future direction of Archival Linked Data development. In both disciplines, engagement with AI to support scholarship and practice is in its infancy. While a detailed examination of current research falls beyond the scope of this article we can note that machine learning techniques have already been explored in a number of Archival Linked Data case-studies, including natural language processing, and its subfield, named entity recognition (Clough et al. 2011; Damova 2020; Pattuelli et al. 2015; Rademaker et al. 2015). It can be supposed that, in automating processes, and making the production of Archival Linked Data less resource intensive, other sub-fields of AI, such as deep learning, computer vision, and neural networks, also offer much potential for the archives sector to scale up the production of Linked Data and innovate how it is accessed. Furthermore, AI has been suggested to offer a potential solution to the challenge of providing controlled access to unstructured data originating from both catalogues and digitised archives (Schreur 2020), and to enable the development of analytical tools meeting the particular and multifaceted requirements of digital humanists requiring access to digital cultural heritage data (Hyvönen 2020b). Drawing on Leopoldina Fortunati's prediction of the increasing "robotization" of the domestic sphere due to the rapid acceptance and incorporation of automated and/or robotic systems into daily life (2018), it is as yet unclear to what extent the datafication of archives will

evolve into the robotisation of archives with the increased incorporation of AI into archival processes.

Regardless, AI cannot be seen as a silver bullet to reducing the resource intensive-ness of the production of Archival Linked Data; there will always be aspects of the production workflow which will not be scalable and human intervention will remain essential. Regardless of how AI is incorporated into the workflow for the production or provision of access to Archival Linked Data, it is essential that any engagement is undertaken critically, with attention directed not only towards its utilisation, but also to its limitations, biases, and ethical implications. Such a conversation is beginning, for example in Jenny Bunn's exploration of engaging with explainable artificial intelligence (XAI) from a recordkeeping perspective (Bunn 2020), and in the examination of four Australian AI case studies to assess the implications of AI for archives (Rolan et al. 2019). Much remains to be addressed, however, and the recent flourishing of archival research networks investigating AI, such as the AEO-LIAN Network (Artificial Intelligence for Cultural Organisations), AURA Network (Archives in the UK/Republic of Ireland and AI), and HAIRA (the Hub for Artificial Intelligence Research in Archives), hints at future progress.

Moss et al. have argued that the move to the digital environment necessitates archive services working collaboratively with other organisations, suggesting web archives, newspaper archives, and data archives as key collaborators (2018). Collaboration is vital for the successful scaling up of the production of Archival Linked Data and the development of tools to make it accessible and interrogable to digital humanists and other users. Furthermore, increasing collaboration across cultural heritage, with other professional sectors and academic disciplines has been widely cited as a key benefit of adopting Linked Data in the archives sector (Llanes-Padrón and Pastor-Sánchez 2017; McKenna et al. 2018; Pitti et al. 2016; vander Sande et al. 2018), with some pointing to increased collaboration as a means of increasing the sustainability of archival Linked Data projects (Martins et al. 2021). Despite evidence of cross-institutional collaboration within Archival Linked Data activity, such collaboration has to date mainly been confined to between multiple cultural heritage institutions and Higher Education Institutes (Smith-Yoshimura 2018). Assessing the success of the Linking Lives project in 2012, Jane Stevenson of the UK Higher Education Archives Hub concluded by pondering: "Maybe we've reached the point in the Linked Data story where we need to focus more strongly on how it will answer the requirements of researchers... Surely we need a more collaborative approach that draws in the technical people, the information professionals and the researchers" (2012b). Even though the value and necessity of engaging with users of archives when developing and introducing new technologies was made clear by the LEADERS project, which centralised users in its early exploration of archival engagement with the World Wide Web (Sexton et al. 2004), it has yet to become commonplace in Archival Linked Data activity.

The Digital Humanities is an inherently interdisciplinary and collaborative field that is information and data-driven, and requires access to large-scale digital datasets. Digital humanists use a range of digital methods, and have a demonstrable interest in, and impetus for, making archival data more digitally accessible. On the other hand, Hannah Lee provided a convincing argument for how archivists, and

others in Library and Information Sciences, could apply their skills and methods to support digital humanities research by better connecting information to its users (2017). There is, therefore, a natural synergy between digital humanists and archival scholars and practitioners which should be exploited to increase the production of, and advocate for investment in, Archival Linked Data. Whereas collaboration between digital humanists and archives professionals to increase the availability of relevant Linked Data datasets would be mutually beneficial in increasing access to digitised archives, cross-sectoral and inter-disciplinary collaboration is also essential to support research on born-digital collections “in order to enable GLAM institutions, institutional networks and infrastructures to develop their born-digital collections in meaningful ways, improve preservation formats, curation workflows, repositories, services and access for researchers” (Ries and Palkó 2019, pp. 4). Calls for such a collaboration are not new—writing in 2013 Matthew Kirschenbaum outlined how archivists and digital humanists need each other: “Digital archivists need digital humanities researchers and subject experts to *use born-digital collections*...Digital humanists need the long-term perspective on data that archivists have...Digital archivists and digital humanists need common and interoperable digital tools...Digital humanists need the collections expertise of digital archivists...Digital archivists need cyberinfrastructure” (2013, para 38). The potential outcomes of such a collaboration are demonstrated by the work of the Shanghai Library which has undertaken a number of digital humanities projects using Linked Data, resulting in the creation of large-scale datasets drawing together archival data from multiple sources. The creation of the Chinese Genealogy Knowledge Service Platform (or Jiapu) is one example which demonstrates how the application of digital humanities methods to digitised archives, in combination with Linked Data, can bring about transformative benefits to users more widely and open up digitised archives to multiple uses: “The Jiapu platform has transformed the ways of accessing genealogy information from providing scanned images to expanding users’ ability to dig deep...The user group has expanded to include scholars for research, the public to find their ancestors, and institutions and developers to create applications by using the linked data from the platform” (Xia and Bao 2020, pp. 81).

Conclusion

Mass investment in digitisation over the past two decades has created a vast resource of digitised archives of varying levels of accessibility. Those which have been subjected to further datafication are more fully accessible for utilisation by digital humanists than those made available only as digital surrogates, and are ripe for further exploitation using digital methods. Linked Data provides a viable means of making digitised and born-digital archives more accessible, producing integrated, enriched, and interoperable large-scale archival datasets available for reuse in multiple ways. To date, the publication of archival metadata as Linked Data has been the key concern of Archival Linked Data activity. However, it is the contents of born-digital and digitised records which offer the greatest potential to researchers, making it increasingly necessary for archive services to datafy their collections. A shared

appreciation of and desire for the provision of access to archives in digital forms and as data suggests that collaboration with digital humanists would be especially fruitful to both the Digital Humanities and the archives sector. Digital humanists could prove a key ally in advocating for Archival Linked Data, demonstrating how it benefits the users of archives, and ensuring that it continues to do so.

The infrastructure required to scale up Archival Linked Data, including standards, best practice, and tools for the creation, maintenance, preservation, and provision of access, are beginning to emerge. Low-barrier, non-sector-specific tools, such as Wikidata, and the use of machine learning and other AI techniques to further automate aspects of the production of Archival Linked Data are on the horizon. Such developments are essential to move beyond individual case studies, expand the sources of archival data being published as Archival Linked Data, and move its production further beyond research institutions in the global north. In the meantime, there are steps that archivists can take on an individual level to prepare themselves and their institutions for engagement with Linked Data when it is more feasible to do so and for progressing their data further up the 5* Linked Open Data Scheme.

The digital environment has forced archival scholars and practitioners to adopt new attitudes to records and archives, viewing them as data, as well as information and records, and to develop a new suite of tools and approaches to process, preserve, and provide access. However, Archival Linked Data requires the development of a new set of technical skills, taking archivists beyond the traditional archival toolbox. It is essential that archivists involve themselves in the production of Archival Linked Data, tools and web services, and engage in scholarly discussion about Linked Data and the Semantic Web, to ensure that the distinguishing characteristics of “archives” are understood and retained. Not only do archivists have unique knowledge of their collections and working practices. Archival scholars and practitioners bring their own set of professional concerns and priorities to the Linked Data environment, ranging from rights management, and the documentation of provenance(s) and other necessary contextual and administrative data, to issues regarding the decolonisation of archival description and descriptive practices, and the long-term preservation of Archival Linked Data. Seth van Hooland and Ruben Verborgh see the growth of archival Linked Data as providing opportunity: “Technology has profoundly influenced the humanities, but we can also leave our marks on technology. Let us hope that the application of linked data principles in our cultural heritage institutions invites people from the humanities and the sciences to pick up the gloves, step into the intellectual boxing ring and engage in a challenging sparring round” (2014, pp. 249). Concerningly, however, Dominic Oldman et al. argue that the lack of context, misrepresentation and ambiguity of historical information published as Linked Data “can be explained, in part, by the lack of engagement or involvement of domain experts themselves in the digital representation of their data, and their lack of knowledge about the possibilities of Semantic technologies, ultimately resulting in the dominance of the technologist at the so-called intersection of digital humanities” (2016, pp. 256). Without the increased involvement and advocacy of archival scholars and practitioners, their professional concerns may continue to be overlooked in the drive to increase access to and datafy digital archives. Their absence is already being felt, diminishing the usability, trustworthiness, and sustainability

of the Archival Linked Data datasets and web services produced, and limiting their potential for utilisation by digital humanists and other users.

Acknowledgements The author wishes to thank Dr Margaret Procter for her comments on an earlier version of this article.

Funding The author's research is funded by the UK Arts and Humanities Research Council and Barclays Group Archives as part of a Collaborative Doctoral Award. In collaboration with Barclays Group Archives, the author was the recipient of a grant from the UK National Archives Archive Testbed Fund in 2020.

Declarations

Conflicts of interest The author has no conflicts of interest to declare that are relevant to the content of this article.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Association of Research Libraries (2019) ARL White paper on wikidata: opportunities and recommendations. <https://www.arl.org/wp-content/uploads/2019/04/2019.04.18-ARL-white-paper-on-Wikidata.pdf>. Accessed 16 Jun. 2021
- Baierer K, Dröge E, Eckert K, Goldfarb D, Iwanowa J, Morbidoni J, Ritze D (2017) DM2E: a linked data source of digitised manuscripts for the digital humanities. *Semant Web* 8(5):733–745
- Balado AB, de Boer V, Schreiber G (2015) Linking Historical Ship Records to a Newspaper Archive. In: Aiello LM, McFarland D (eds) *Social Informatics: SocInfo 2014 Workshops*, Barcelona, Spain, November 11, 2014, Revised Selected Papers. Springer, Cham, pp 254–263
- Berners-Lee T (2006) Design Issues: Linked Data <https://www.w3.org/DesignIssues/LinkedData.html>. Accessed 20 Jun. 2021
- Berners-Lee T (2010) Design issues: is your linked open data 5 Star <https://www.w3.org/DesignIssues/LinkedData.html>. Accessed 20 Jun. 2021
- Bikakis A, Hyvönen E, Jean S, Markhoff B, Mosca A (2021) Editorial: special issue on semantic web for cultural heritage. *Semant Web* 12(2):163–167
- Bizer C, Heath T, Berners-Lee T (2009) Linked data – the story so far. *Int J Semant Web Inf Syst* 5(3):1–22
- Blanke J, Riechert T (2020) Towards an RDF knowledge graph of scholars from early modern history. In: 2020 IEEE 14th international conference on semantic computing, 3–5 February 2020, San Diego, California. IEEE Conference Publishing Services, Los Alamitos, pp 471–72. <https://doi.org/10.1109/ICSC47212.2020.9309131>.
- Blanke T, Prescott A (2016) Dealing with Big Data. In: Griffin G, Hayler M (eds) *Research methods for reading digital data in the digital humanities*. Edinburgh University Press, Edinburgh, pp 184–205
- Bones H (2019) Linked digital archives and the historical publishing world: an Australasian perspective. *Hist Compass* 17(3):e12522. <https://doi.org/10.1111/hic3.12522>

- Browell G (2016) From linked open data to linked open knowledge. In: Baker D, Evans W (eds) *Digital strategies: from applications and content to libraries and people*. Elsevier, London, pp 87–99
- Bruneau O, Lasolle N, Lieber J, Nauer E, Pavlova S, Rollet L (2021) Applying and developing semantic web technologies for exploiting a corpus in history of science: the case study of the Henri Poincaré correspondence. *Semant Web* 12(2):359–378
- Bunn J (2020) Working in contexts for which transparency is important: a recordkeeping view of explainable artificial intelligence (XAI). *Rec Manag J* 30(2):143–153
- Chiarcos C, Pagé-Perron E, Khait I, Schenk N, and Reckling L (2018) Towards a Linked Open Data Edition of Sumerian Corpora. In: Calzolari N, Choukri K, Cieri C, Declerck T, Goggi S, Hasida K, Isahara H, et al. (eds) *Proceedings of the 11th International Conference on Language Resources and Evaluation*, May 2018, Miyazaki, Japan. European Language Resources Association, Paris, pp 2437–44
- Cimiano P, Chiarcos C, McCrae JP, Garcia J (2020) *Linguistic linked data: representation, generation and applications*. Springer, Cham
- Clough P, Tang J, Hall MM, Warner A (2011) Linking archival data to location: a case study at the UK national archives. *Aslib Proc* 63(2–3):127–147. <https://doi.org/10.1108/00012531111135628>
- Corn A, Patrick SWJ (2019) Exploring the applicability of the semantic web for discovering and navigating Australian indigenous knowledge resources. *Arch Manuscr* 47(1):131–152. <https://doi.org/10.1080/01576895.2019.1575248>
- Damova M (2020) Linked open data prototype of the historical archive of the European commission. In: Nault J (ed) *Archiving 2020 Online: Digitization, Preservation, and Access - Final Program and Proceedings*, 7–21 May 2020, online. Society for Imaging Science and Technology, Springfield, pp 92–97. <https://doi.org/10.2352/issn.2168-3204.2020.1.0.92>
- Daquino M, Mambelli F, Peroni S, Tomasi F, Vitali F (2017) Enhancing semantic expressivity in the cultural heritage domain: exposing the Zeri photo archive as linked open data. *J Comput Cult Heritage* 10(4):1–21. <https://doi.org/10.1145/3051487>
- Davis K (2019) Old metadata in a new world: standardizing the Getty Provenance Index for linked data. *Art Libr J* 44(4):162–166. <https://doi.org/10.1017/alj.2019.24>
- de Boer V, Oomen J, Inel O, Aroyo L, van Staveren E, Helmich W, de Beurs D (2015) DIVE into the event-based browsing of linked historical media. *J Web Semant.* <https://doi.org/10.1016/j.websem.2015.06.003>
- Debruyne C, Beyan OD, Grant R, Collins S, Decker S, Harrower N (2016) A semantic architecture for preserving and interpreting the information contained in Irish historical vital records. *Int J Digit Libr* 17:159–174. <https://doi.org/10.1007/s00799-016-0180-8>
- Dobreski B, Park J, Leathers A, Qin J (2019) Remodeling archival metadata descriptions for linked archives. In: Golub K, Zeng M, Walk P, Oh S, Baker T (eds) *Proceedings of the international conference on Dublin Core and metadata applications*, 23–25 September 2019, Seoul, South Korea. Dublin Core Initiative, Online, pp 1–11
- Egloff M, Picca D (2020) WeDH - A Friendly Tool for Building Literary Corpora Enriched with Encyclopedic Metadata. In: Calzolari N, Béchet F, Blache P, Choukri K, Cieri C, Declerck T, Goggi S, et al. (eds) *Proceedings of the 12th international conference on language resources and evaluation May 2020*, Marseille. European Language Resources Association, Paris, pp. 813–16
- Eyharabide V, Lully V, Morel F (2019) MusicKG: Representations of Sound and Music in the Middle Ages as Linked Open Data. In: Acosta M, Cudré-Mauroux P, Maleshkova M, Pellegrini T, Sack H, Sure-Vetter Y (eds) *Semantic systems. The power of AI and knowledge graphs. SEMANTiCS 2019*, 9–12 September 2019, Karlsruhe, Germany. Springer, Cham, pp 57–63. https://doi.org/10.1007/978-3-030-33220-4_5
- Fortunati F (2018) Robotization and the domestic sphere. *New Media Soc* 20(2):2673–2690
- García EGB, Manaiescu M, Ros S (2016) From syllables, lines and stanzas to linked open data: standardization, interoperability and multilingual challenges for digital humanities. In: García-Peñalvo FJ (ed) *TEEM'16: proceedings of the fourth international conference on technological ecosystems for enhancing multicultural diversity 2–4 November 2016*, Salamanca, Spain. Association for Computing Machinery, New York, pp 979–83. <https://doi.org/10.1145/3012430.3012635>
- Gartner R (2015) An XML schema for enhancing the semantic interoperability of archival description. *Arch Sci* 15:295–313. <https://doi.org/10.1007/s10502-014-9225-1>
- Godby CJ, Wang S, Mixter JK (2015) Library linked data in the cloud: OCLC's experiments with new models of resource description. Morgan & Claypool, California

- Gould T (2018) Using the hive mind: wikidata integration and artist Pages. <https://medium.com/nationalgalleries-digital/using-the-hive-mind-wikidata-integration-and-artist-pages-618574fa9168>. Accessed on 16 Jun. 2021
- Gracy KF (2015) Archival description and linked data: a preliminary study of opportunities and implementation challenges. *Arch Sci* 15:239–294. <https://doi.org/10.1007/s10502-014-9216-2>
- Gracy KF (2018) Enriching and enhancing moving images with linked data: an exploration in the alignment of metadata models. *J Doc* 74(2):354–371. <https://doi.org/10.1108/JD-07-2017-0106>
- Gracy KF, Zeng ML, Skirvin L (2013) Exploring methods to improve access to music resources by aligning library data with linked data: a report of methodologies and preliminary findings. *J Am Soc Info Sci Technol* 64(10):2078–2099. <https://doi.org/10.1002/asi.22914>
- Herbert D, Parilla L (2021) Linking the Enslaved at Dogue Run Farm, paper presented at LD4 2021, July 20 2021, <https://www.youtube.com/watch?v=he5d4erB9wA>. Accessed 20 Sep. 2021
- Hoekstra R, Meroño-Peñuela A, Dentler K, Rijpma A, Zijdemann R, Zandhuis I (2016) An ecosystem for linked humanities data. In: Sack H, Rizzo G, Steinmetz N, Mladenčić D, Auer S, Lange C (eds) *The semantic web. ESWC 2016 Satellite Events, May 29-June 2 2016, Heraklion, Crete, Greece*. LNCS vol 9989. Springer, Cham, pp. 425–440. https://doi.org/10.1007/978-3-319-47602-5_54
- Horne R (2020) Beyond lists: digital gazetteers and digital history. *Historian* 82(1):37–50. <https://doi.org/10.1080/00182370.2020.1725992>
- Hunter J, Macarthur J, van der Plaats D, Gosseye J, Wilson A, Muys A, Bannerman G (2012) A web 3.0 Approach to building an online digital archive of architectural practice in post-war Queensland. *Comma*. <https://doi.org/10.3828/comma.2012.2.5>
- Hyvönen E (2012) Publishing and Using Cultural Heritage Linked Data on the Semantic Web. Morgan & Claypool, California
- Hyvönen E (2020a) Linked open data infrastructure for digital humanities in Finland. In: Reinsone S, Skadina I, Baklāne A, Daugavietis J (eds) *Proceedings of the digital humanities in the Nordic Countries 5th Conference, 21–23 October 2020, online*. CEUR, Aachen, pp 254–259
- Hyvönen E (2020b) Using the semantic web in digital humanities: shift from data-publishing to data-analysis and serendipitous discovery. *Semant Web* 11(1):187–193
- Hyvönen E, Heino E, Leskinen P, Ikkala E, Koho M, Tamper M, Tuominen J, Mäkelä E (2016) War-Sampo Data Service and Semantic Portal for Publishing Linked Open Data about the Second World War History. In: Sack H, Blomqvist E, d’Aquin M, Ghidini C, Ponzetto S, Lange C (eds) *The Semantic Web. Latest Advances and New Domains. ESWC 2016, May 29-June 2 2016, Heraklion, Crete, Greece*. LNCS vol 9678. Springer, Cham, pp 758–773. https://doi.org/10.1007/978-3-319-34129-3_46
- Hyvönen E, Leskinen P, Tamper M, Rantala H, Ikkala E, Tuominen J, Keravuori K (2019) BiographySampo – publishing and enriching biographies on the semantic web for digital humanities research. In: (eds) *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, LNCS vol 11503. Springer, Cham, pp. 574–89
- International Federation of Library Associations and Institutions (2017) Opportunities for academic and research libraries and wikipedia. <https://www.ifla.org/files/assets/hq/topics/info-society/iflawikipeiaopportunitiesforacademicandresearchlibraries.pdf>. Accessed 16 Jun. 2021
- Isaac A, Waites W, Young J, Zeng M (2011) W3C Library Linked data incubator group: datasets, value vocabularies, and metadata element sets. <https://www.w3.org/2005/Incubator/ldd/XGR-ldd-vocab-dataset-20111025/>. Accessed 17 Jun. 2021
- Jaillant L (2019) After the digital revolution: working with emails and born-digital records in literary and publishers’ archives. *Arch Manuscri* 47(3):285–304. <https://doi.org/10.1080/01576895.2019.1640555>
- Jordanous A, Hedges M, Lawrence KF, Tupman C (2012) Exploring manuscripts: sharing ancient wisdoms across the semantic web. In: Burdescu DD, Akerkar R, Bădică C (eds) *WIMS ’12: Proceedings of the 2nd international conference on web intelligence, mining and semantics, 13–15 June 2012, Craiova, Romania*. Association for Computing Machinery, New York, pp 44–55
- Kapsalis E (2019) Wikidata: recruiting the crowd to power access to digital archives. *J Radio Audio Media* 26(1):134–142. <https://doi.org/10.1080/19376529.2019.1559520>
- Keller MA, Parsons J, Glaser H, Calter M (2011) Report of the Stanford linked data workshop, 27 June–1 July 2011. Council on library and information resources, Alexandria, Virginia. Accessed 25 Jun. 2021 at: <https://www.clir.org/pubs/reports/pub152/stanford-linked-data-workshop/>

- Kirschenbaum M (2013) The .txtual condition: digital humanities, born-digital archives, and the future literary. *Digit Humanit* Q 7(1): Art 6
- Koho M, Ikkala E, Leskinen P, Tamper M, Tuominen J, Hyvönen E (2021) WarSampo knowledge graph: Finland in the second world war as linked open data. *Semant Web* 12(2):265–278. <https://doi.org/10.3233/sw-200392>
- Koho M, Hyvönen E, Erkki H, Jouni T, Leskinen P, Mäkelä E (2016) Linked death – representing, publishing, and using Second World War death records as linked open data. In: Adamou A, Daga E, Isaksen L (eds) *The semantic web: ESWC 2017 satellite events. ESWC 2017. Lecture notes in computer science*, vol 10577, 28 May–1 June 2017, Portorož, Slovenia. Springer, Cham, pp 369–383. https://doi.org/10.1007/978-3-319-70407-4_45
- Koster L, Woutersen-Windhouwer S (2018) FAIR Principles for Library, Archive and Museum Collections: A proposal for standards for reusable collections. *Code4Lib Journal* 40. <https://journal.code4lib.org/articles/13427>. Accessed 20 Jun. 2021
- Lee H (2017) Is it Worth It? The library and information science degree in the digital humanities. In: Hai-Jew S (ed) *Data analytics in digital humanities multimedia systems and applications*. Springer, Cham, pp 275–295
- Llanes-Padrón D, Pastor-Sánchez JA (2017) Records in contexts: the road of archives to semantic interoperability. *Program* 51(4):387–405. <https://doi.org/10.1108/PROG-03-2017-0021>
- Martins P, Costa L, Ramalho JC (2021) Knowledge graph of press clippings referring social minorities. In: Lopes CT, Ribiero C, Niccolucci F, Rodrigues I, Freire N (eds) *Linked archives international workshop working notes*, pp 40–47. available at: <https://linkedarchives.inesctec.pt/>. Accessed 5 Oct. 2021
- Mayer-Schönberger V, Cukier K (2013) *Big data: a revolution that will transform how we live, work, and think*. John Murray Publishers, London
- McKenna L, Debruyne C, O’Sullivan D (2018) Understanding the Position of Information Professionals with Regards to Linked Data: A Survey of Libraries, Archives and Museums. In: Chen J, Gonçalves MA, Allen JM (eds) *JCDL ’18: Proceedings of the ACM/IEEE Joint Conference on Digital Libraries*, 3–7 June 2018, Fort Worth, Texas, USA. Association for Computing Machinery, New York, pp 7–16. <https://doi.org/10.1145/3197026.3197041>
- McKenna L (2020) Benefits of interlinking in digital archives, paper presented at AURA workshop 1, November 17 2020, <https://www.youtube.com/watch?v=CcvCj7aLqhg&t=8s>. Accessed 8 Jun. 2021
- Miyakita G, Leskinen P, Hyvönen E (2018) Using linked data for prosopographical research of historical persons: case U.S. Congress Legislators. In: Ioannides M, Fink E, Brumana R, Patias P, Doulamis A (eds) *Digital heritage. Progress in cultural heritage: documentation, preservation, and protection*. Springer, Cham, pp 150–62
- Morris RJ (2017) Document to database and spreadsheet. In: Gunn S, Faire L (eds) *Research methods for history*, 2nd edn. Edinburgh University Press, Edinburgh, pp 147–169
- Moss M, Thomas D, Gollins T (2018) The reconfiguration of the archive as data to be mined. *Archivaria* 86:118–151
- Neubert J (2019) 20th Century Press archives: data donation to wikidata. <http://zbw.eu/labs/en/blog/20th-century-press-archives-data-donation-to-wikidata>. Accessed 25 Jun 2021
- Niu J (2016) Linked data for archives. *Archivaria* 82(December):83–110
- Noble SU (2018) Critical surveillance literacy in social media: interrogating black death and dying online. *Black Camera Int Film J* 9(2):147–160. <https://doi.org/10.2979/blackcamera.9.2.10>
- Oldman D, Doerr M, Gradmann S (2016) Zen and the art of linked data: new strategies for a semantic web of humanist knowledge. In: Schreibman S, Siemens R, Unsworth J (eds) *A new companion to digital humanities*, 2nd edn. Wiley Blackwell, New Jersey, pp 251–273
- Open Knowledge Foundation (ND) (2021) Open Definition. <https://opendefinition.org/>. Accessed 18 Jun. 2021
- Padilla T (2018) Collections as data: implications for enclosure. *Coll and Res Libr News* 79(6):296–300
- Pattuelli MC, Miller M, Lange L, Fitzell S, Li-Madeo C (2013) Crafting linked open data for cultural heritage: mapping and curation tools for the linked jazz project. *Code4Lib Journal* 21. <https://journal.code4lib.org/articles/8670>. Accessed 20 Jun. 2021
- Pattuelli MC, Provo A, Thorsen H (2015) Ontology building for linked open data: a pragmatic perspective. *J Libr Metadata* 15(3–4):265–293. <https://doi.org/10.1080/19386389.2015.1099979>

- Pattueli MC, Hwang K, Miller M (2017) Accidental discovery, intentional inquiry: leveraging linked data to uncover the women of jazz. *Digit Scholarsh Humanit* 32(4):918–924. <https://doi.org/10.1093/llc/fqw047>
- Pitti D, Stocking B, Clavaud F (2016) An introduction to “Records in Contexts”: an archival description draft standard. *Comma* 1–2:173–188
- Proffitt M (2020) Libraries think they are notable: using WorldCat holdings data to fill gaps in Wikipedia. <https://hangingtogether.org/?p=8222>. Accessed 16 Jun. 2021
- Program for Cooperative Cataloguing (2018) PCC strategic directions: January 2018–December 2021. <https://www.loc.gov/aba/pcc/about/PCC-Strategic-Directions-2018-2021.pdf>. Accessed 16 Jun. 2021
- Rademaker A, Oliveira DAB, de Paiva V, Higuchi S, MedeirosSá A, Alvim M (2015) A linked open data architecture for the historical archives of the Getúlio Vargas foundation. *Int J Digit Libr* 15(2–4):153–167. <https://doi.org/10.1007/s00799-015-0147-1>
- Ranade S, (2016) Traces through time: a probabilistic approach to connected archival data. In: Proceedings of the 2016 IEEE international conference on big data, 5–8 December 2016, Washington, DC, USA. IEEE pp 3260–3265. <https://doi.org/10.1109/BigData.2016.7840983>
- Rekkavik A (2014) RDF linked data cataloguing at oslo public library. *SCATnews: newsletter of the standing committee of the IFLA Cataloguing Section* 41: 13–16
- Ries T, Palkó G (2019) Born-digital archives. *Int J Digit Humanit* 1:1–11
- Robledano-Arillo J, Navarro-Bonilla D, Cerdá-Díaz J (2019) Application of linked open data to the coding and dissemination of Spanish Civil War photographic archives. *J Doc* 76(1):67–95. <https://doi.org/10.1108/JD-06-2019-0112>
- Rolan G, Humphries G, Jeffrey L, Samaras E, Antsouпова T, Stuart K (2019) More human than human? Artificial intelligence in the archive. *Arch Manuscri* 47(2):179–203. <https://doi.org/10.1080/01576895.2018.1502088>
- Romein CA, Kemman M, Birkholz JM, Baker J, De Gruijter M, Meroño-Peñuela A, Ries T, Ros R, Scagliola S (2020) State of the field: digital history. *Hist* 105(365):291–312
- Romeo F (2016) Bringing [art] knowledge to everyone who seeks it. <https://medium.com/digital-moma/bringing-art-knowledge-to-everyone-who-seeks-it-899ec257a55c>. Accessed 16 Jun. 2021
- Ruddock B (2011) Linked data and the LOCAH project. *Bus Info Rev* 28(2):105–111. <https://doi.org/10.1177/0266382111404013>
- Schneider J, Adams C, DeBauche S, Echols R, McKean C, Moran J, Waugh D (2019) Appraising, processing, and providing access to email in contemporary literary archives. *Arch Manuscri* 47(3):305–326. <https://doi.org/10.1080/01576895.2019.1622138>
- Schreibman S, Siemens R, Unsworth J (2004) The digital humanities and humanities computing: an introduction. In: Schreibman S, Siemens R, Unsworth J (eds) *A companion to digital humanities*. Blackwell Publishing, Oxford, pp xxiii–xxvii
- Schreur PE (2020) The use of linked data and artificial elements as key elements in the transformation of technical services. *Cat Class Q* 58(5):473–485. <https://doi.org/10.1080/01639374.2020.1772434>
- Schüll ND (2020) Self-Tracking. In: Thylstrup NB, Agostinho D, Ring A, D’Ignazio C, Veel K (eds) *Uncertain archives: critical keywords for big data*. The MIT Press, Cambridge, Massachusetts, pp 457–468
- Sexton A, Turner C, Yeo G, Hockey S (2004) Understanding users: a prerequisite for developing new technologies. *J Soc Arch* 25(1):33–49. <https://doi.org/10.1080/0037981042000199133>
- Smith-Yoshimura K (2018) Analysis of 2018 International Linked Data Survey for Implementers. *Code4Lib Journal* 42. <https://journal.code4lib.org/articles/13867>. Accessed 20 Jun. 2021
- Southerton C (2020) Datafication. In: Schintler L, McNeely C (eds) *Encyclopaedia of big data*. Springer, Cham
- Stevenson J (2012a) Linking lives: creating an end-user interface using linked data. *Inf Stands Q* 24(2–3):14–23
- Stevenson J (2012b) Linking lives evaluation report. <http://linkinglives.archiveshub.ac.uk/2012/11/15/linking-lives-evaluation-report/>. Accessed 18 Jun 2021
- Summers E, Isaac A, Redding C, Krech D (2008) LCSH, SKOS, and Linked Data. In: DC-2008: proceedings of the international conference on dublin core and metadata applications, 22–26 September 2008, Berlin, Germany. Dublin Core Initiative, Online, pp 25–33
- Sutherland T (2019) The carceral archive: documentary records, narrative construction, and predictive risk assessment. *J Cult Anal* 10(22148/16):039

- Sztyler T, Huber J, Noessner J, Murdock J, Allen C, Niepert M (2014) LOD: Linking digital humanities content to the web of data. In: Klein M, Rauber A, Cunningham SJ (eds) Proceedings of the 2014 IEEE/ACM joint conference on digital libraries 8–12 September 2014, London, UK. IEEE, pp 423–424 <https://doi.org/10.1109/JCDL.2014.6970206>
- Tamper M, Leskinen P, Apajalahti K, Hyvönen E (2018) Using biographical texts as linked data for prosopographical research and applications. In: Ioannides M, Fink E, Brumana R, Patias P, Doulamis A, Martins J, Wallace M (eds) Digital heritage. Progress in cultural heritage: documentation, preservation, and protection. EuroMed 2018. LNCS vol 11196, 29 October – 29 November 2018, Nicosia, Cyprus. Springer, Cham pp 125–37. https://doi.org/10.1007/978-3-030-01762-0_11
- Thornton K (2017) Library of Congress Digital Preservation Using Wikidata URIs! <https://campuspress.yale.edu/knowledgepermaculture/2017/02/27/library-of-congress-digital-preservation-using-wikidata-uris/>. Accessed 16 Jun 2021
- Tillman RK (2016) Opportunities for encoding EAD for linked data extraction and publication. *J Arch Organ* 13(1–2):19–36. <https://doi.org/10.1080/15332748.2017.1400725>
- van Hooland S, Verborgh R (2014) Linked data for libraries, archives and museums: how to clean, link and publish your metadata. Facet Publishing, London
- vander Sande M, Verborgh R, Hochstenbach P, van de Sompel H (2018) Toward sustainable publishing and querying of distributed linked data archives. *J Doc* 74(1):195–222. <https://doi.org/10.1108/JD-03-2017-0040>
- Wikipedia (2021) Wikipedia: Notability (people). [https://en.wikipedia.org/wiki/Wikipedia:Notability_\(people\)](https://en.wikipedia.org/wiki/Wikipedia:Notability_(people)). Accessed 12 Jun. 2021
- Wilkinson MD, Dumontier M, Aalbersberg I et al (2016) The FAIR guiding principles for scientific data management and stewardship. *Sci Data*. <https://doi.org/10.1038/sdata.2016.18>
- Winters J, Prescott A (2019) Negotiating the born-digital: a problem of search. *Arch Manuscri* 47(3):391–403. <https://doi.org/10.1080/01576895.2019.1640753>
- Xia C, Liu W, Zhang L (2018) Implementation of a linked data-based genealogy knowledge service platform for digital humanities. *Data Info Manag* 2(1):15–26
- Xia C, Bao X (2020) Dynamic digital humanities projects from shanghai library in China. In: McAleer M, Deyrup MM (eds) Transformative digital humanities: challenges and opportunities. Routledge, Abingdon, Chap Six
- Zeng ML (2019) Semantic enrichment for enhancing LAM data and supporting digital humanities. review article. *El Prof De La Inf* 28(1):1–35. <https://doi.org/10.3145/epi.2019.ene.03>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ashleigh Hawkins is a doctoral research candidate at the Liverpool University Centre for Archive Studies, funded by the UK Arts and Humanities Research Council and Barclays Group Archives. Her research interests lie in archives and technology (specifically Linked Data and AI), the Digital Humanities, and the history of recordkeeping. In collaboration with Barclays Group Archives, she was awarded a grant from the UK National Archives 'Archives Testbed Fund' in 2020 to undertake a practical case study exploring the application of Linked Data to increase access to historic customer records.