Open Library of Humanities

Photo Archives and Linked Open Data. The Added Value

Marilena Daquino, Department of Classical Philology and Italian Studies, University of Bologna, Italy, Marilena.daquino2@unibo.it

In the last two decades, Cultural Heritage institutions have been revisiting the way they publish their data. Due to the rise of Semantic Web technologies and graph-based search engines, the shift in the technology stack has required many to reconsider the way their data is organised. The appreciable byproduct of this phenomenon has been the development of data literacy skills among cataloguers, archivists, and collection managers, who were in turn promised a revamp of the archival institution's image in terms of authoritativeness (due to the improved data quality) and attractiveness towards patrons (due to the enhanced search capabilities). In this article we describe how photo archives have embraced such a new paradigm, and we discuss benefits and limitations, moving from a representative example, i.e., ZERI & LODe, a project devoted to the publication of the catalogue of the Federico Zeri Photo Archive into Linked Open Data. The focus of the analysis is the added value promised by Semantic Web technologies and the Open Data business model to cataloguers, scholars, and arts enthusiasts.

Introduction

The rise of Semantic Web technologies has required organisations in several knowledge domains to revise their technology stack and knowledge organisation arrangements. In particular, the Cultural Heritage domain has been one of the most enthusiastic adopters since the beginning of the Linked Open Data (LOD) movement (Berners-Lee, Hendler and Lassilla, 2001), and to date it represents one of its major application fields (Bikakis et al., 2021).

Libraries, archives, and museums (LAMs) have widely embraced the new paradigm, although with diverse degrees of adoption. Notably, scholarship in Cultural Analytics and Digital Humanities is replete with case studies testifying how institutions have successfully moved into the realm of Semantic Web. Typically, such projects address some technical challenges arisen by legacy data and software solutions, and present a new data source, tool, or framework to tackle such (research) problems. The ultimate goal is often achieving the benefits promised by the Semantic Web, such as improved information retrieval, facilitated record linking, improved visibility, and better analytics and services/applications.

However, despite such benefits being consistently referenced as the motivating factor for moving into the Semantic Web, we cannot find extensive discussion and evaluation of those achievements in the literature (Hawkins, 2022). In fact, (1) while projects demonstrate that they have reached the technological goal, they do not share evidence that benefits have been achieved, nor to what extent; (2) despite the promise of better analytics and new services, traditional research methods are still being used in online catalogues; (3) small institutions that rely on aggregators to publish their data cannot afford to develop tools tailored to their collections. In this context, photo archives offer a representative, despite understudied, example. Photo archives have increasingly made their collections available via digital catalogues to support scholars in iconographic and historiographic research (Robledano-Arillo, Navarro-Bonilla and Cerdá-Díaz, 2020). Particularly, art historical photo archives are attracted by LOD as a way to address the lack of structured metadata about images content and as a *lingua franca* to integrate photo collections across institutes, with the goal to support scholars with advanced search capabilities (Daquino, 2019; Delmas-Glass and Sanderson, 2020; Caraffa et al., 2020).

The objective of this article is to investigate how photo archives have embraced Semantic Web technologies and whether expectations have been fulfilled, in terms of long-term results and acquired skills. To guide the analysis, we present the ZERI & LODe project as an example of LOD of a small although renowned art historical photo archive that has been running long enough to allow us to discuss achievements

and limitations. In particular, we are interested in understanding what the added value of LOD is for photo archives and whether frustrating elements may be hiding behind unfulfilled promises. While we do not claim that all the conclusions in the domain/example under examination can be generalised to LAMs, we believe some considerations are of general interest and would deserve further investigation also in other fields.

The remainder of the article is the following. In the section *Related work* we present a summary of benefits brought in the Cultural Heritage domain by Semantic Web technologies, as described in articles in Computer Science and Digital Humanities fields. In section *Photo Archives and Linked Open Data*, we describe the landscape of Photo Archival Linked Open Data, describing the degree of their adoption and the limitations they have faced. In section *The Zeri & LODe project* we describe our case study. Sections *The added value and the barriers* address benefits and limits brought by the technology shift in an attempt to generalise conclusions derived from the case study.

Related work

In the last twenty years, several works in the Computer Science and Digital Humanities scholarship have highlighted benefits of the Semantic Web in the Cultural Heritage domain (Mitchell, 2016; McKenna, 2018). Notably, over the years, such positive expectations appear to be more and more tailored to the requirements of the Cultural Heritage domain, which has in turn became a champion of those technologies. Expectations can be summarised into four main promises as discussed below, namely: improved information retrieval, facilitated data integration and enrichment, decentralisation and improved visibility, better analytics and services/applications (McKenna et al., 2018).

- **Improved information retrieval.** Based on the assumption that humanities studies are interested in relations (or semantic paths) between artefacts, events, people, places, etc., the Semantic Web would overtake traditional keyword-based approaches to retrieve information, which ignore the meaning (hence failing in concept disambiguation) and the interrelations that concepts have, and would foster smarter search applications (Benjamins et al., 2004; Lodi et al., 2017)
- **Facilitated data integration and enrichment.** Based on the assumption that publishing Cultural Heritage data on the Web stimulates cultural tourism, creative economy, and collaborations between institutions, data must be syntactically and semantically interoperable, which is ensured by the (consistent) usage of, respectively, the RDF model and domain ontologies (Hyvonen, 2022).

- **Decentralisation and improved visibility**. On the one hand, Linked Open Data allows anyone to publish data about anything (even about objects preserved by other institutions) anywhere on the Web, hence fostering a decentralised approach to publish, and later access, information on the Web. On the other hand, semantic interoperability allows small and medium institutions that do not have resources or expertise to share their data as LOD to do so via national or international aggregators, such as Europeana, hence improving institutional visibility and discoverability of resources (De Boer et al., 2012).
- **Better analytics and services/applications**. Seamlessly integrating data sources to perform analysis or populating mashup applications are intriguing possibilities offered by technical/semantic interoperability. The extensive cleaning work performed ahead to create LOD significantly simplifies data wrangling and harmonisation tasks, which are usually time-consuming preliminary research activities (Davis, 2019; Hawkins, 2022).

While nobody naively claims that Semantic Web technologies are the panacea to all the problems, it has been argued that they offer a tool set to solve issues more effectively (Hyvonen, 2022). Cultural Heritage and archival LOD have been widely recognised as beneficial to scholars in the Digital Humanities research field (Llanes-Padrón and Pastor-Sánchez, 2017; Daquino et al., 2017; McKenna et al., 2018; Robledano-Arillo, Navarro-Bonilla and Cerdá-Díaz, 2020; Giagnolini et al., 2023).

However, a clear assessment of such promises is often missing (Hawkins, 2022) and there is no evidence that such benefits are also appreciated by other stakeholders, e.g. cataloguers, scholars in other domains, industry, or lay users. Some scholars in Computer Science have criticised the feasibility of the premises of Semantic Web themselves, and a significant number of scholars in the Semantic Web community believe the original vision has not been realised yet (Hogan, 2020). Although Knowledge Graphs are increasingly adopted in industrial use cases due to their demonstrated or perceived added value, no formal evaluation of their benefits seems to be available (Hitzler, 2021). Surveys with Information Professionals (McKenna et al., 2018) demonstrate that cataloguers and archivists recognise potentialities of the Semantic Web but struggle to fully enjoy such benefits due to technical barriers. Evaluations of Semantic-Web-based interfaces with lay users have been a disregarded matter for a long time (Hawkins, 2021), since Semantic Web has often targeted niche groups (i.e. engineers and scientists) (Hachey and Gasevic, 2011). To the best of our knowledge, there are no general evaluations carried out with lay users on the perception of benefits derived from the technology adoption.

In this article we contribute to the debate by providing an assessment of aforementioned promises with respect to Photo Archival LOD and how these affect cataloguers and lay users.

Photo Archives and Linked Open Data

Archives and photo archives have adopted different strategies to embrace Semantic Web technologies, mainly due to the different nature of resources they describe.

Archives generally create archival records of fonds, series, and folders, and rarely include detailed descriptions of individual documents. To this extent, they adopt ontologies and vocabularies based on descriptive fields available in archival standards, e.g. Records in Contexts (EGAD, 2019). Archives cooperate in consortia devoted to the publication of partial information, such as SNAC (Larson et al., 2014), which publishes EAC-CPF records of people and organisations found in archival collections, or Europeana, which publishes a subset of metadata of archival records. Moreover, some archives have published their collection data individually.

The Italian *Istituto per i beni artistici culturali e naturali* (IBC) has been among the first institutions experimenting with ontologies for representing archival records and has contributed to the development of software solutions for browsing and exploring graph data (Mazzini and Ricci, 2011). Over the years the project was renamed ReLOAD, and several new, selected collections have been transformed and integrated. The goal of the project is to experience the benefits of LOD in terms of improved accessibility for final users (citizens, institutions, and companies), facilitating the development of new applications. Unfortunately, such expectations are still presented as future endeavours (Ricci, 2017).

The LOCAH project, later revamped in the Linked Lives project (LOCAH and Stevenson, 2012; Browell, 2015) has been a pioneer in producing the LOD catalogue of the UK Archive Hub. The follow-up project was moved by the urge of showing the benefits provided by the new technologies to final users of the archives. The collaboration with the SNAC project generated a number of visualisations of people's archival records. Being an early attempt to experiment with LOD, fixing data and technological issues was the core of the activities, and the evaluation of achievements has been postponed.

The National Archives of the United Kingdom (Garmendia and Retter, 2021) have moved their databases into a pan-archival LOD catalogue based on the Records in Contexts Conceptual Model (RiC-CM) and a combination of vocabularies inspired by The Matterhorn RDF Model (Dubois and Wildi, 2019). One of the main advantages of such a change—in the view of cataloguers—is the possibility to effectively represent,

store, and retrieve provenance and versioning information of their records, which was not possible with legacy technologies. However, while legacy data have been successfully transformed into RDF, existing cataloguing and user interfaces have not been abandoned yet. Expectations regard (1) future cost savings, by replacing existing legacy and unsupported software, reducing duplication, (2) creating new opportunities through unlocking the unrealised potential in data (Garmendia and Retter 2021), and (3) the linkage to external resources (such as Legislation.gov.uk, Office for National Statistics, government datasets, and Wikidata).

The Archives Nationales of France (ANF) developed a reusable tool to convert EAD finding aids and EAC-CPF authority records into RDF files according to RiC-O (Francart et al., 2021). However, only a selection of data is available on a GitHub repository, and no interface is provided for querying the data. The ALEGORIA research project has made available a RDF/RiC-O dataset derived from the collections of aerial photographs preserved at the ANF. Again, data are released as static files, but specialised multi-modal search engines are built on top of photographs for iconographic research and use metadata to provide context information.

It is rather common that academic research projects take over in the transformation and publication of archival data on behalf of institutions, which do not always have the means to integrate LOD catalogues into their current workflows or cannot afford to redesign and replace their user interfaces (Daquino, 2021). Other examples include experiments in knowledge extraction and knowledge graph generation starting from the full-text of archival documents, photographs, or metadata records, such as the EPISA project on the Portuguese National Archives (Varagnolo et al., 2021; Koch et al., 2023), which extracted events and (exceptionally) produced a graph according to CIDOC-CRM. The Major Minors project is another national project, where information about social minorities are extracted from press clippings of Portuguese newspapers (Martins, Costa and Ramalho, 2021). The ARTchives project aims at collecting archival descriptions of art historians' archives and describe them using the Wikidata model. Scholars involved in the project have experimented with data mining and relation extraction methods in order to develop recommendation systems for historians (Giagnolini et al., 2023). Unfortunately, most results, regardless of these being individual projects or collaborative efforts, are still in a prototypical phase.

Similarly to archives, photo archives provide descriptions of the hierarchical structure of their collections, but also include detailed information of single photographs and their subjects. In this respect, photo archives tend to adopt standards closer to libraries and museums, where the focus is on the "social biography" of the artefact (Gosden, Larson and Petch, 2007). Photo archives contribute to collaborative projects too. In

2016 the Europeana project reported that it had digitised over 48 million photographs (Schneider and Weinberg, 2020), which are described according to the Europeana Data Model and provide basic metadata to a broad audience.

Since 2013, 14 art historical photo archives have been actively collaborating in the PHAROS consortium to publish the wealth of their data collections and make it accessible via a bespoke integrated platform (Caraffa et al., 2020). The online platform (Binkowski, 2022), based on a customisation of Research Space (Oldman and Tanase, 2018) gathers about three out of 20 million images belonging to archives across Europe and North America (Binkowsky, 2023), which agreed on leveraging museum vocabularies and ontologies, such as CIDOC-CRM (Le Boeuf et al., 2016), and Getty vocabularies (Harpring, 2010), and ICONCLASS (Brandhorst and Posthumus, 2016). Since the subject of documentary photographs are artworks, and artwork metadata have priority when satisfying their patrons' enquiries, PHAROS partners archives actively collaborate with the Linked Art project (Delmas-Glass and Sanderson, 2020), which gathers expertise from several museums around the world to define a shared data model for artwork description. Moreover, the project actively experiments novel methods for image similarity, therefore facilitating cataloguing and matching tasks across archives and dissemination via IIIF (Klic, 2023). While a few photo archives have also individually published their collections as Linked Open Data, e.g. the Zeri Photo Archive (Daquino et al., 2017), the Getty Research Collections,[1] and Bernard Berenson's catalogue *The drawings of the Florentine painters* (Klic et al., 2017), most partners rely on the PHAROS infrastructure to share a LOD catalogue separately from their traditional collection management systems.

Another notable example is the work done on the photographic archives from the Swiss Society for Folklore Studies (SSFS) as part of the PIA project (Cornut, Raemy and Spiess, 2023). Photographs metadata have been transformed into RDF/CIDOC-CRM, again reusing the Linked Art data model and IIIF standard. The newly created collections are published using OmekaS platform[2] and computer vision methods are applied to annotate photographs.

Examples of individual photo archives publishing their data as LOD are limited. The Siberian SB RAS Photographic Archive (Krayneva and Marchuk, 2020) created its own ontology-based platform, called SORAN 1957, to serve about 24,000 scans of photographs. The Spanish Civil War photographic archives (Robledano-Arillo, Navarro-Bonilla and Cerdá-Díaz, 2020) have developed an ontology for describing

---

[1] https://www.getty.edu/research/collections/, last accessed 20 August 2024.

[2] https://omeka.org/s/, last accessed 20 August 2024.

their catalogue and produced a sample dataset to validate it, but no working prototypes are available to users. The Linked Stage Graph project has transformed data about 7000 black and white photographs from the National Archive of Baden-Wuerttemberg about the Stuttgart State Theatre into RDF according to another bespoke ontology (Tietz et al., 2023). Data are accessible via a dedicated Web application and two visualisation tools, i.e. LODview and Vikus viewer.

The Zeri & LODe Project

The Zeri & LODe project is a pilot project to transform a subset of the Federico Zeri Photo Archive catalogue into LOD (Daquino et al., 2017). The art historical photo archive is a member of the PHAROS consortium and it experimented with Semantic Web technologies in early stages, developing a prototype of ontologies, two mapping documents, datasets, and interlinking options relevant to other partners. Developed assets and services are the following:

- Two ontologies, respectively called F Entry Ontology and OA Entry Ontology, which are mostly based on CIDOC-CRM, PROV-O, and the SPAR ontologies. The ontologies allow one to describe the structure of the archive, individual photographic documents, depicted artworks, attribution ship, artwork provenance, bibliography (i.e., the library of Federico Zeri), people and organisations involved in the objects' life cycle and their role.
- Two mapping documents to respectively address terminological aspects and alignment of metadata standards used by the Zeri photo archive (i.e. the photograph metadata set, the artwork metadata set, the authority files of artists, photographers, and auction catalogues) into CIDOC-CRM terms.
- A RDF dataset, published online[3] and served via a dedicated platform for querying (via SPARQL endpoint) and browsing (via LODview). URIs of photographs and artworks are linked and accessible from the current Zeri online catalogue records, so as to allow a smooth transition between the legacy catalogue and the RDF browsing experience. Records also include links to several authorities (ULAN, VIAF, Wikidata, geonames, ICONCLASS, AAT). Versioned copies of the dataset are available in the institutional repository for long-term preservation.

Peculiarities and similarities between the Zeri & LODe project and projects described above are several, namely:

---

[3] http://data.fondazionezeri.unibo.it, last accessed 20 August 2024.

- **Non-native LOD catalogue.** The project created a non-native LOD catalogue, which lives a separate life from the legacy catalogue, with its own interfaces and life cycle. This is mostly the result of a request of the representatives of the archive, who were not ready to replace the current cataloguing system with neither a LOD-native cataloguing software nor a semi-static data catalogue.

- **Academic project.** A team of digital humanists, computer scientists, and domain experts contributed to the realisation of the prototype, which therefore has a strong research imprint. The prototype has been running for more than 8 years, hosted by the University of Bologna, and so far, it has requested one significant intervention for updating and migrating the software infrastructure to a new machine. The maintenance is granted by the Digital Humanities Advanced Research Centre of the University of Bologna, which ensures the long-term preservation of the data and the services developed.

- **Research focuses on conceptual aspects.** Scholars involved in the project co-designed with archivists and art historians bespoke ontologies to address peculiarities of the archival data that were not representable with existing ontologies, hence focusing the research on conceptual, descriptive, aspects. Relevant new aspects included relations between people (e.g. influence), artworks (e.g. copies), and artefacts of different nature (e.g. citations, distribution of images). The usage of CIDOC-CRM as a building block became immediately evident when figuring future works would be devoted to the dissemination of the dataset among art historians, who are the most significant target audience of the art historical photo archive. To describe the archive and the photographic object, standards from the publishing domain were reused instead.

- **Limited services for data dissemination.** The development of a limited number of services for disseminating RDF data (i.e. a RDF browser and a SPARQL endpoint) is due to the need of minimising the expenses for maintenance, hence ensuring a sustainable solution over time. Moreover, archive personnel had mixed feelings towards alternative interfaces to their current catalogue, which they thought may distract users rather than attract new ones.

Like other projects, the Zeri & LODe project was moved by the promises of the Semantic Web (Daquino et al. 2017), which revealed being an attractive solution for a number of reasons, namely:

- **Improving the quality of the cataloguing data.** Expensive data cleansing and normalisation operations have been performed to extract clean data to be

transformed into RDF. The Zeri archive staff is trained in using editorial rules consistently, therefore metadata extraction methods did not require extensive revisions. However, limitations of prior metadata standards obliged cataloguers to "improperly" use some metadata fields to record more than one piece of information (hence the need for clear editorial rules to handle such situations). For instance, the field dedicated to the reason supporting an artwork attribution included both a controlled term (e.g. "bibliography") and a reference to relevant documents (e.g. a bibliographic in-text reference). The usage of LOD and the possibility to design their own new data model allowed cataloguers to free the potential of such hidden pieces of information and make them searchable.

· **The perspective of record alignment.** Currently Wikidata, IBC, and the Ministry of Italian Cultural Heritage (MIC), have included explicit links to the Zeri data. However, interlinking addresses entities like people, places, and organisations, does not include an alignment between cultural objects (e.g. the artworks depicted in photographs). Therefore, the reconciliation is only superficial and does not really allow a seamless transition between datasets, nor does it allow any institution to enrich their data by automatically importing significant data from aligned sources. To be effective, interlinking must be performed between data sources that present overlapping information, such as other PHAROS members, where photographs of the same artworks (and sometimes the very same photographs) are preserved in more than one institute, or museum catalogues that include detailed information on the artworks depicted in the photographs. This would allow performing researches across institutional collections that are not currently possible.

· **The increased visibility.** Zeri data and images are available in the PHAROS research platform, Europeana, and CulturaItalia, and the archive is described in ARTchives. Such aggregators contribute to increase the visibility of the institution, since cataloguing records can be accessed via several—unpredictable—entry points on the Web. While it is not possible to confirm such a claim using information collected by the aggregators (e.g. user analytics in the aforementioned platforms), we collect user analytics on the usage of Zeri data. Analytics show us that around 40% of user views come from external sources, while 60% of users instead come from links in the current online catalogue. Moreover, the collaboration in consortia allows the archive to be more visible in institutional networks. Collaborations foster credibility in the eye of (1) funders,

hence increasing chances of getting fundings, (2) other institutions, encouraging them in participating in collaborative projects, and (3) stakeholders, proposing themselves as reliable innovation leaders.

- **Empowering patrons and users of open data**. The increased visibility on the Web allows archives to attract scholars, developers, and companies that are interested in accessing and reusing available open data in creative applications. To the best of our knowledge, the Zeri & LODe project was the driver of one PhD thesis, five master theses, ~10 scholarly publications in international venues for Cultural Heritage, Digital Humanities, and Semantic Web communities (~100 citations), two follow up projects (Daquino, 2019; Giagnolini et al., 2023), and it is currently listed by the Ministry of Italian Cultural Heritage as a gold standard and prototype to be imitated for creating the new national digital library. Currently the LOD catalogue is also used in Digital Humanities master courses as teaching material to learn methods for data visualisation and data analysis. Around five student projects leverage the dataset in websites presenting data storytelling journeys. However, proactive users of the LOD catalogue do not include art historians, who lack the technical skills to manipulate the data and perform quantitative art history research. To this extent, historians are limited to the legacy search interfaces offered by the institution.

Notice that among the benefits does not appear the need of changing legacy technologies, which is rather perceived as an obstacle. On the one hand, the archive uses a cataloguing system developed by a Web agency which does not allow exporting data, therefore hampering a smooth transition to other software solutions, and includes custom solutions tailored on the information system desired by the archivists, which are difficult to reproduce in new solutions. On the other hand, integrating the LOD catalogue in the current system is cumbersome and not viable, since it would require extensive revision.

Lastly, cataloguers' personal growth and acquisition of data literacy skills can be considered a nice byproduct of the project. The knowledge transfer process that occurred between digital humanists, computer scientists and archive personnel, allowed the latter to continue pursuing interdisciplinary research (e.g., Giagnolini et al., 2023) and actively collaborate in new projects that make extensive use of Semantic Web technologies, such as PHAROS. Likewise, the possibility to use the Zeri dataset for teaching purposes and to explain the complexity of documents interconnections in

the Cultural Heritage ecosystem by means of a notable example, has been a precious opportunity for the new generation of Digital Humanists, who can better appreciate the articulations of quantitative art history and history of photography.

The Added Value and the Barriers

From the overall picture outlined in the previous sections, it appears clear that the promises of the Semantic Web are very attractive to institutions that struggle to serve structured complex information to their patrons and would like to allow them to pursue sophisticated research via intuitive interfaces.

In this respect, data integration across institutions and record linking seems to be an appreciated feature enabled by Linked Open Data, since they effectively contribute to accomplish the mission of cultural institutions, i.e. supporting patrons in knowledge discovery (LOCAH and Stevenson, 2012; Ricci, 2017; Garmendia and Retter, 2021). In particular, data integration opens to new opportunities in the development of information retrieval and analytical tools—which would ideally leverage information coming from different data sources—and it compensates for data quality issues that inevitably affect individual institutions, merging (partial) information belonging to multiple sources. To this extent, institutions appreciate that the expensive work in data cleansing required to perform the alignment is a necessary cumbersome activity that prevents users from doing it manually by themselves, hence preventing an important element of frustration in data reuse.

Projects seem also to invite an undefined audience of developers, stakeholders, and citizens in reusing their open data creatively, developing applications and performing studies that would not be possible with legacy technologies. However, the literature does not show many examples of such projects where lay people, humanists, and representatives of Cultural Heritage institutions autonomously manipulate LOD for their (research) purposes. Instead, multidisciplinary teams are always needed, projects are mostly developed in academia, and they require resources to be pursued and later maintained. We can then assume that the wide range of opportunities offered by the (linked) open data business model is accessible only by a minority of tech-savvy people, who graciously support humanists in understanding and reframing their research questions using quantitative methods, and managing their expectations in terms of results. A common aspect characterising scholarly projects is that these tend to be dismissed once the research trigger is lost, and long-term maintenance cannot be ensured.

In fact, the landscape sketched above shows that most of such projects are still in a prototypical phase and the advanced applications enabled by LOD are yet to

be developed. For instance, student projects mostly address a subset of data and aim at answering one or more rather specific research questions, and they do so via data analysis and visualisation. However, results of such projects are shared as stories, blog posts, or websites where the reader is not allowed to manipulate and filter data used for the analysis and can only passively appreciate the message intended by the "story tellers". Barriers are also posed by legacy technology, data quality issues, and the highly encouraged conformance to standard and popular ontologies, which seem to hamper the development of sophisticated solutions for disseminating data.

Despite Semantic Web technologies having been around for twenty years, the debate on how to reuse ontologies for describing the Cultural Heritage is still open, and different approaches are in place. Efforts often take the form of communities (e.g. PHAROS and Linked Art), where members agree to compromise in order to achieve the great goal of data integration (Daquino et al., 2017; Delmas-Glass and Sanderson, 2020; Koch et al., 2023; Cornut, Raemy and Spiess, 2023). In many other cases, small-size projects decide to develop their own ontologies (Daquino et al., 2017; Dubois and Wildi, 2019; Krayneva and Marchuk, 2020; Robledano-Arillo, Navarro-Bonilla and Cerdá-Díaz, 2020; Tietz et al., 2023), so as not to compromise data quality, be able to manage changes in the ontology, and to speed up the project development. Nonetheless, the never-ending discussion on ontological aspects has increased the awareness on a topic overlooked before the advent of Semantic Web, that is, the description and publication of provenance information as a way to promote trust in data users and to provide valuable insights into record-keeping behaviours (Garmendia and Retter, 2021). Unfortunately, the scattered landscape of ontology reuse practices affects such a topic too, and it is a significant barrier to the settlement of Semantic Web technologies as everyday practice in cultural institutions.

As a consequence, the majority of projects have set the publication of their data as Linked Open Data as an immediate milestone, postponing advanced applications that would effectively make value out of data to follow-up projects. In some cases, institutions that could not afford the transformation of their data have delegated this task to Cultural Heritage aggregators (e.g. Europeana, PHAROS, CulturaItalia), therefore delegating also the development of ontologies and applications to leverage such data. In both cases (individual or aggregate publishing), resulting LOD catalogues are usually new, separate assets that live separately from the original catalogues, often creating misalignment of data sources in small institutes (De Boer et al., 2012). Moreover, it has been argued that aggregators are not designed to support a wide range of user informative needs (Peroni, Tomasi and Vitali, 2013).

Conclusion

In summary, it seems clear since the very early stages of Semantic Web adoption in archives (LOCAH and Stevenson, 2012) that *just* publishing Linked Data is not enough to reach the promised benefits, while plenty of work has to be done to showcase how data are to be used, and to empower a community of data reusers that goes beyond the privileged group of digital humanists and computer scientists.

Research in the last years has been focusing more on the development of reusable tools that simplify the creation of Linked Open Data (Daquino et al., 2023; Oldman and Tanase, 2018), as well as to visualise and narrate the added value of such data (Renda et al., 2023). While a few solutions have reached informal consensus among Cultural Heritage institutions (e.g. LODview), scholars acknowledge the lack of satisfying means to leverage Linked Open Data without having a solid knowledge of technological aspects (Hawkins, 2021; Chen, 2023) and complain about the lack of generous interfaces (Whitelaw, 2015) that would allow serendipitous discovery and would create a more inclusive environment for citizens and lay users. More generally, it has been argued that there is still little knowledge about users' needs (Hawkins, 2021).

Considering recent advances in AI technologies (knowledge graphs, deep learning, automated knowledge base construction, language models, computer vision, and multimodality), we can expect that many of the challenges here presented will be tackled (if not solved) using more powerful and effective means (Alam et al., 2023), envisioning a future where the usage of cultural AI is free of the technology legacy burden, and energies could be spent more productively in creative applications.

Acknowledgements

Competing Interests

References

Alam, M, Boer, VD, Daga, E, Erp, MV, Hyvönen, E and Meroño-Peñuela, A 2023 Editorial of Special Issue on Cultural Heritage and Semantic Web Technology. *Semantic Web–Interoperability, Usability, Applicability*, 14(2): 1–4.

Benjamins, VR, Contreras, J, Blázquez, M, Dodero, JM, Garcia, A, Navas, E, Hernández, F and Wert, C 2004 Cultural Heritage and the Semantic Web. In: *European Semantic Web Symposium*. Berlin, Heidelberg: Springer Berlin Heidelberg. pp. 433–444.

Berners-Lee, T, Hendler, J and Lassila, O 2001 The Semantic Web. *Scientific American*, 284(5): 34–43.

Bikakis, A, Hyvönen, E, Jean, S, Markhoff, B and Mosca, A 2021 Special Issue on Semantic Web for Cultural Heritage. *Semantic Web*, 12(2): 163–167.

Binkowski, K 2022. PHAROS opens its online portal. *Art Libraries Journal*, 48(1): 15–20.

Brandhorst, H and Posthumus, E 2016 Iconclass: a key to collaboration in the digital humanities. In: *The Routledge Companion to Medieval Iconography*. Routledge. pp. 201–218.

Browell, G 2015 From linked open data to linked open knowledge. In: Baker, D. and Evans, W (eds.) *Digital information strategies: From applications and content to libraries and people*. Chandos Publishing. pp.87–99.

Caraffa, C, Pugh, E, Stuber, T and Ruby, LW 2020 PHAROS: A digital research space for photo archives. *Art Libraries Journal*, 45(1): 2–11.

Chen, YN 2023 An investigation of linked data catalogue features in libraries, archives, and museums: a checklist approach. *The Electronic Library*, 41(5): 700–721.

Cornut, M, Raemy, JA and Spiess, F 2023 Annotations as Knowledge Practices in Image Archives: Application of Linked Open Usable Data and Machine Learning. *ACM Journal on Computing and Cultural Heritage*, 16(4): 1–19.

Davis, K 2019 Old metadata in a new world: Standardizing the Getty Provenance Index for linked data. *Art Libraries Journal*, 44(4): 162–166.

Daquino, M 2019 *Mining Authoritativeness in Art Historical Photo Archives: Semantic Web Applications for Connoisseurship*. IOS Press.

Daquino, M 2021 Linked Open Data native cataloguing and archival description. *JLIS.it*, 12(3): 91–104.

Daquino, M, Mambelli, F, Peroni, S, Tomasi, F and Vitali, F 2017 Enhancing semantic expressivity in the Cultural Heritage domain: exposing the Zeri Photo Archive as Linked Open Data. *ACM Journal on Computing and Cultural Heritage*, 10(4): 1–21.

Daquino, M, Wigham, M, Daga, E, Giagnolini, L and Tomasi, F 2023 CLEF. A linked open data native system for crowdsourcing. *ACM Journal on Computing and Cultural Heritage*, 16(3): 1–17.

De Boer, V, Wielemaker, J, Van Gent, J, Hildebrand, M, Isaac, A, Van Ossenbruggen, J and Schreiber, G 2012 Supporting linked data production for Cultural Heritage institutes: the Amsterdam museum case study. In: T*he Semantic Web: Research and Applications: 9th Extended Semantic Web Conference, ESWC 2012, Heraklion, Crete, Greece, May 27–31, 2012*. Berlin, Heidelberg: Springer Berlin Heidelberg. pp. 733–747.

Delmas-Glass, E and Sanderson, R 2020 Fostering a community of PHAROS scholars through the adoption of open standards. *Art Libraries Journal*, 45(1). pp.19–23.

Dubois, A and Wildi, T 2019 The Matterhorn RDF Data Model [sides] iPRES. https://www.alaarchivos.org/wp-content/uploads/2017/12/3.-Alain-Dubois-Andreas-Nef.pdf [Last Accessed 19 January 2024].

EGAD 2019 International Council on Archives Records in Contexts Ontology (ICA RiC-O) version 0.1. https://www.ica.org/standards/RiC/RiC-O_v0-2.html [Last Accessed 19 January 2024].

Francart, T, Clavaud, F and Charbonnier, P 2021 RiC-O converter: a software to convert EAC-CPF and EAD 2002 XML files to RDF datasets conforming to records in contexts ontology. In: *Proceedings of the Linked Archives International Workshop*. pp. 30–36.

Garmendia J and Retter A 2021 Developing a Pan-Archival Linked Data Catalogue. In: Berget, G, Hall, MM., Brenn, D and Kumpulainen, S *Linking Theory and Practice of Digital Libraries*. Springer International Publishing. pp. 93–103

Giagnolini, L, Daquino, M, Mambelli, F and Tomasi, F 2023 Exploratory methods for relation discovery in archival data. *Digital Scholarship in the Humanities*, 38(1): 111–126.

Gosden, C, Larson, F and Petch, A 2007 *Knowing Things: exploring the collections at the Pitt Rivers Museum 1884–1945*. Oxford University Press.

Hachey, G and Gasevic, D 2011 Semantic Web user interfaces: A systematic mapping study. Athabasca University. Semantic Web.

Harpring, P 2010 Development of the Getty vocabularies: AAT, TGN, ULAN, and CONA. *Art Documentation: Journal of the Art Libraries Society of North America*, 29(1): 67–72.

Hawkins, A 2021 Advocating for linked archives: the benefits to users of archival linked data. In: Berget, G, Hall, MM., Brenn, D and Kumpulainen, S *Linking Theory and Practice of Digital Libraries*. Springer International Publishing. pp. 52–63.

Hawkins, A 2022 Archives, linked data and the digital humanities: increasing access to digitised and born-digital archives via the semantic Web. *Archival Science*, 22(3): 319–344.

Hitzler, P 2021 A review of the semantic Web field. *Communications of the ACM*, 64(2): 76–83.

Hogan, A 2020 The semantic Web: Two decades on. *Semantic Web*, 11(1): 169–185.

Hyvonen, E 2022 *Publishing and using Cultural Heritage linked data on the semantic Web*. Springer Nature.

Klic, L 2023 Linked Open Images: Visual similarity for the Semantic Web. *Semantic Web*, 14(2): 197–208.

Klic, L, Miller, M, Nelson, JK, Pattuelli, CM and Provo, A 2017 The drawings of the Florentine painters: From print catalog to Linked Open Data. *The Code4Lib Journal*, 38.

Koch, I, Teixeira Lopes, C and Ribeiro, C 2023 Moving from ISAD (G) to a CIDOC CRM-based Linked Data Model in the Portuguese Archives. *ACM Journal on Computing and Cultural Heritage*, 16(4): 1–21.

Krayneva, I and Marchuk, A 2020 Open Archives of the SB RAS: Systems of Historical Factography. In: *Proceedings of the 22nd Conference on Scientific Services & Internet (SSI-2020)*. pp. 189–200.

Larson, RR, Pitti, D and Turner, A 2014 SNAC: The Social Networks and Archival Context project – Towards an archival authority cooperative. In: *IEEE/ACM joint conference on digital libraries*. pp. 427–428.

Le Boeuf, P, Doerr, M, Ore, CE, and Stead, S 2016 Definition of the CIDOC Conceptual Reference Model. Technical Report 6.2. https://www.cidoc-crm.org/sites/default/files/2018-10-26%23CIDOC%20CRM_v6.2.4_esIP.pdf [Last Accessed 19 January 2024].

Llanes-Padrón, D and Pastor-Sánchez, JA 2017 Records in contexts: the road of archives to semantic interoperability. *Program*, 51(4): 387–405.

Lodi, G, Asprino, L, Nuzzolese, AG, Presutti, V, Gangemi, A, Recupero, DR, Veninata, C and Orsini, A 2017 Semantic Web for Cultural Heritage valorisation. In: Hai-Jew, S (eds) *Data Analytics in Digital Humanities*. Multimedia Systems and Applications. Springer, Cham. https://doi.org/10.1007/978-3-319-54499-1_1

Martins, P, Costa, L, Ramalho, JC 2021 Knowledge Graph of Press Clippings Referring Social Minorities. In: Berget, G, Hall, MM., Brenn, D and Kumpulainen, S *Linking Theory and Practice of Digital Libraries*. Springer International Publishing.

Mazzini, S and Ricci, F 2011 EAC-CPF Ontology and Linked Archival Data. In *SDA*. pp. 72–81.

McKenna, L, Debruyne, C and O'Sullivan, D 2018 Understanding the position of information professionals with regards to linked data: a survey of libraries, archives and museums. In: *Proceedings of the 18th ACM/IEEE on Joint Conference on Digital Libraries*. pp. 7–16.

Mitchell, ET 2016 Library linked data: early activity and development. *ALA TechSource*, 52(1).

Oldman, D and Tanase, D 2018 Reshaping the knowledge graph by connecting researchers, data and practices in ResearchSpace. In: *The Semantic Web–ISWC 2018: Proceedings of the 17th International Semantic Web Conference, Monterey, CA, USA, October 8–12, 2018*. pp. 325–340.

Peroni, S, Tomasi, F and Vitali, F 2013 Reflecting on the europeana data model. In: *Digital Libraries and Archives: 8th Italian Research Conference, IRCDL 2012, Bari, Italy, February 9–10, 2012, Revised Selected Papers*. pp. 228–240.

Renda, G, Daquino, M and Presutti, V 2023 Melody: A Platform for Linked Open Data Visualisation and Curated Storytelling. In: *HT '23: Proceedings of the 34th ACM Conference on Hypertext and Socia Media*. pp.1–8. https://doi.org/10.1145/3603163.3609035

Robledano-Arillo, J, Navarro-Bonilla, D and Cerdá-Díaz, J 2020 Application of Linked Open Data to the coding and dissemination of Spanish Civil War photographic archives. *Journal of Documentation*, 76(1): 67–95.

Ricci F 2017 Accesso libero. I linked open data pubblicati da IBC intervento della giornata di incontro IBC con gli operatori bibliotecari e archivistici. Presented at *Giornata di incontro IBC con gli operatori bibliotecari e archivistici*, 17 gennaio 2017 – Bologna. https://online.ibc.regione.emilia-romagna.it/I/libri/pdf/biblioteche-archivi-2017/ricci.pdf [Last Accessed 19 January 2024].

Schneider, J and Weinberg, P 2020 No Way Back–Reflections on the Future of the African Photographic Archive. *History in Africa*, 47: 167–194.

Tietz, T, Bruns, O and Sack, H 2023 A Data Model for Linked Stage Graph and the Historical Performing Arts Domain. In: *Proc. of the International Workshop on Semantic Web and Ontology Design for Cultural Heritage (SWODCH)*.

Varagnolo, D, Rodrigues, C, Martins, A, Melo, D, and Rodrigues I 2021 Extracting Entities and Events from Archives Textual Metadata. In: Berget, G, Hall, MM., Brenn, D and Kumpulainen, S *Linking Theory and Practice of Digital Libraries*. Springer International Publishing.

Whitelaw, M 2015 Generous interfaces for digital cultural collections. *Digital Humanities Quarterly*, 9(1).