
Abstract

Artist websites are sites of art where ‘digital social memory’ gets accrued over time. As such, they require preservation care similar to other forms of cultural heritage. This essay describes several sets of medium-specific, socio-technical challenges to archiving artist websites connected to the processes of both archiving published materials on the web, as well as subsequently providing access to these materials via an indexed database. These challenges can be grouped under the rubrics of ethics, technology, curation, as well as cataloguing and sustainability. The primary focus of the essay is to discuss the challenges – and opportunities – associated with cataloguing and structuring digital archive (meta)data, by drawing upon findings from the practice-led research project to redesign the ArtBase archive of net art, alongside two further case studies. All case studies present critical engagements with Linked Open Data (LOD) software environments as means to collaboratively organize, store and provide access to heterogeneous archive data on the web. The possibility to capture and represent complex networks of relations and dependencies through the flexible structure of LOD is key to the concluding policy recommendations.

Author

Lozana Rossenova is a digital designer and researcher, focusing on open-source and community-driven approaches to digital archives. In 2021, she completed a PhD at the Centre for the Study of the Networked Image (London South Bank University) in collaboration with Rhizome, a leading international born-digital art organisation. Rossenova is currently based at the Open Science Lab at TIB, Hanover, working on the NFDI4Culture project towards a national research infrastructure of cultural heritage data.

Art Doc Archive is developing a prototype for the centralized archiving of the materials self-documented by Berlin-based artists on websites and in social media. The “Reclaim your Archive!” series of articles accompanies the project and is edited by Bianca Ludewig. All texts can be found on the project’s blog: reclaim.hypotheses.org

Art Doc Archive was funded by:
Sites of art and networks of data: Archiving artists’ websites as linked open data

Lozana Rossenova

Introduction

Publishing on the web – texts, various media files and other ‘raw’ data – has become a staple activity across nearly all domains of contemporary culture. It is a use-case well served by a range of software tools that are readily accessible to users with a variety of skill sets. The ease with which individual users can purchase a piece of server space and maintain their own website(s) or simply publish materials to platforms hosted by third parties (whether corporations, or non-profits) may seem trivial nowadays, yet the promise of permanence remains elusive. The danger of losing access to previously published material is always present: link rot, unpaid server bills, outdated PHP versions, Silicon Valley tycoons buying/selling and sunsetting platforms. At the same time, the use-case for archiving web ‘publications’, and subsequently maintaining databases that allow us to index and search remains largely underserved. Web-archiving efforts at the national and international level (e.g., IIPC, Internet Archive) cannot address the needs of individual cultural producers (Rossenova, 2017a), whereas software for publishing ‘small’ databases with indexed materials on the web remains largely in the purview of passionate experts rather than the general public (Critchlow, 2023).

In this context, the task of archiving artists’ websites acknowledges their role as important sites of our collective digital social memory (Connor, 2016; Rossenova, 2017b) but faces the socio-technical challenges of both archiving published materials on the web as well subsequently providing access to these materials via an indexed database. In this essay, I will draw out some parallel concerns and showcase relevant developments based on findings from my own PhD research (carried out 2016–2021) and from several related professional-practice projects.

The PhD project sought to better understand problems associated with the archiving of net art and to address the question of how the artworks can be made accessible to the public in their native environment – online – while enabling users of the archive to gain an expanded understanding of what constitutes the art object. Net art works are not single digital objects, but rather assemblages (Dekker, 2018), dependent on specific software/hardware environments to be executed and rendered. They often change over time and require specific user input in order to be performed. In this regard they share much in common with various other forms of artistic use of the web, including use for documentation and publication purposes. Like net art works, many artists’ websites feature diverse orchestrations of software components, media files, and require user interaction. Interactions can be scripted in predictable or surprising ways – all part of the artists’ intent. Capturing and preserving that shares a range of challenges also present in the ArtBase archive, the main object of study of my PhD research. The ArtBase was established in
1999 as an online space to showcase and archive net art. It is maintained by Rhizome, an international art organisation championing born-digital art and culture through primarily online-based exhibitions, special events and Research & Development activities focused on preservation.

This essay considers artist websites as sites of art, which may document, present and publish artworks and information about artistic activities, or be considered artworks in their own right. In all cases, archiving these websites requires attention and care around technological obsolescence, cultural context and intricate networks of relations – between servers and data; between artists and their works – also observed in the findings from redesigning the ArtBase archive (Rossenova, 2021).

Challenges for archiving artists’ websites

Drawing on the experience of redesigning the ArtBase archive, the challenges associated with archiving artists’ websites, or any form of artists’ publishing on the web can be grouped in the following five focus areas:

- **Ethical** – primarily encompassing privacy concerns (how will sensitive data be handled by the archive?, who will have access?, etc.); notions of consent (will the archive acquire consent from all concerned parties for the archiving activities?); as well as the right to be forgotten (i.e., observing intentional decisions by the creators not to preserve/archive personal content);

- **Technical** – concerning the preservation of user interactivity (i.e., archiving code and data, and not just static videos and media); software dependencies (i.e., client- and server-side software changes over time); and third-party data dependencies (i.e., concerning the archiving of external data sources, code libraries, media storage that may affect the performativity of the original site).

- **Curatorial** – concerning the selection of materials to be archived and the identification of the curatorial boundaries around the website – both of these aspects are also connected to technical challenges regarding external dependencies; the curatorial decision whether to follow external links, how to deal with embedded materials – and especially materials published via social media – have ethical dimensions (i.e., would consent from others also be required?) as well as technical (i.e., would the archiving of a third-party site bring in additional technical challenges?).

- **Cataloguing** – concerning the choice of a taxonomy to organize the index of archived materials and how/if the index links out to other cataloguing databases; this set of challenges also extends to the choice of software to organize the index and the type of data architecture supported by this software – e.g., relational vs. graph database – which brings its own sets of limitations and opportunities when it comes to archival (meta)data description.

- **Sustainability** – concerning the long term responsibility for keeping and maintaining the archive; this set of challenges is interconnected with all other focus points, since it encompasses: a) the ethical responsibility of keeping up with potential updates to artistic intentions, b) the technical aspects of obsolescence and updating dependencies, c) the curatorial challenge of who is responsible for decision making and what policies guide the decision-makers, and d) the cataloguing question of how the published materials are described and what ‘status’ in the canon of artistic production do they hold, which
ultimately informs what type of cultural institution needs to take responsibility over long
term preservation and maintenance.

It is beyond the scope of this essay to discuss in depth all the different challenges outlined above.
Instead, relevant sources are referenced in the bibliography, and one primary area of concern is
discussed in more detail below – namely, the question of cataloguing the archival index, which in
many ways can also illustrate the entangled socio-technical issues that play a role across all sets
of challenges. Conventional CMSs (collection management systems), which rely on the closed
architecture of hierarchical databases with a limited set of predefined standard categories, limit
the possibility to describe archival objects that fall outside canonical forms, and furthermore
create siloes, which means it is not possible to easily link out to and connect knowledge across
institutions and collections that may hold related objects. Meeting the challenges of cataloguing
archives of artists’ websites with alternative software and data-protocol solutions aims to show
how digital cultural heritage needs to and can be guided by alternative curatorial and institutional
policies when it comes to materials published on the web.

The potential of linked open data to describe networked objects

Adopting Linked Open Data (LOD) for cultural heritage institutions means publishing their digital
collections in a machine-readable, openly-licensed, networked environment, rather than siloed
webpages (Delmas-Glass and Sanderson, 2020). LOD is based on the concept of the Semantic Web
(Berners-Lee, et al., 2001), in which machine-readable ‘facts’ or ‘claims’ are organized using web
addresses (i.e., URLs or IRIs). In this manner, heterogeneous vocabularies and data sources can
coexist in the same space, just as all websites are accessible via the same addressing schema.

LOD has been a popular concept among GLAMs (Galleries, Libraries, Archives and Museums) for
at least a decade now, but wide-spread implementation is yet to be achieved (Rossenova, 2021).
Amid institutions tied to legacy systems and endemic shrinking of resources for the arts, the
majority of commercial systems that offer LOD infrastructure are not a viable option. But focusing
only on the difficulty of switching over from legacy systems to new infrastructure, or on the high
learning curve and oftentimes high costs associated with running a standards-compliant LOD
system for collections management, misses out on some of the key opportunities with this
alternative way of structuring data.

Considering artists’ websites and other artists’ materials published on the web as vital parts of
our digital cultural heritage and, hence, materials important to be preserved in the context of
traditional memory institutions (i.e., GLAMs) puts new pressures on traditional archival and
object description standards as well as collection-management software implementations. These
standards and software systems already struggle to address justified concerns in classification
bias based on race, gender and outdated colonial knowledge structures (Nesmith, 2006; Shilton
and Srinivasan, 2007; Christen, 2012). Adding to that pressure is the notion that an artists’
website or web publication cannot simply be described as a digital surrogate for an artwork, or a
piece of documentation. As evidenced in the example of the ArtBase archive, net art assemblages,
their technical dependencies and the relations between artists, contributors and users were too
complex to describe following standard metadata terms and vocabularies (Rossenova, 2020).
Artists’ websites similarly encompass a range of assembled materials – raw data, texts and media;
depend on various technical infrastructures; and can manifest complex relations between artists, galleries, exhibitions and more. LOD and its basic principle of organising information in the form of semantic triples offers a great amount of flexibility to describe complex networks of relations, to introduce new categories of objects and to make it all still possible to search and index, and crucially – interlink within a growing web of semantic data (the Linked Open Data cloud). The web of semantic data makes further information about people, places, institutions, as well as abstract concepts such as genres and movements, themes and knowledge domains, available for search and interlinking with one’s own data, without the need to hold redundant copies of all that data. While few institutions have their complete collections in the LOD cloud, many institutions do share data on artists and collection objects in Wikidata (Stinson, 2017), the world’s largest public knowledge base that is part of the LOD cloud. Wikidata and its underlying software – Wikibase – offer new possibilities for archives of digital cultural heritage. As open source software solutions including graphical user interfaces, they have the potential to make LOD accessible to a broad set of cultural organisations and less formal cultural initiatives (Fauconnier, et al., 2018; Rossenova, 2021). In turn, these organisations and initiatives can adopt new policies and practices in terms of how they describe the materials in their care and more fundamentally – what materials as a whole they are able to collect and preserve.

**Wikidata and Wikibase as sites for enacting networked relations**

Wikidata and Wikibase are two software packages from the Wikimedia family of applications. They allow the storage and management of structured machine-readable data, in addition to retaining common characteristics of wikis, such as collaboration and version control features. Wikidata is the public-facing instance of the software – a type of Wikipedia for structured data. Wikibase is the open-source software environment built to run Wikidata, released in 2015 by Wikimedia Germany (a chapter of the Wikimedia Foundation).

Being a free, collaborative and multilingual online space, with a capacity to collect structured data via a graphical user interface, Wikidata lowers the barrier for the general public to work with, contribute and develop further the W3C’s vision for a web of LOD (Wikidata, 2022). There have been a number of successful collaborative projects between the Wikimedia Foundation and GLAM (Galleries, Libraries, Archives and Museums) institutions focusing on Wikidata’s potential role as an aggregator, or a hub, of centralized cultural data connecting various vocabularies across institutions and languages, and leveraging the contributions of volunteers from the general public (Stinson, 2017; Europeana, 2017).

Since all Wikimedia software is open source and has public domain licensing, Wikibase can also be deployed as a separate instance. As an independent software system, Wikibase offers most of the technical benefits of Wikidata, as well as some further possibilities for customisation. Rhizome is one of the first cultural heritage organisations to pilot the use of Wikibase as an internal knowledge management system and explore more fully the potential of structuring data that spans artworks, artists, technical dependencies and events using the ‘open world’ data model of Wikibase’ LOD implementation (Fauconnier, et al., 2018). That is, a data model that lacks any pre-set metadata categories and relations, and instead offers the possibility to build a completely custom system for describing born-digital cultural objects and establishing machine-readable links across entities in the database and the wider LOD cloud. What is more, using a ‘wiki’
software with a build-in account management and version control system means it is possible to do this process iteratively and collaboratively – involving different stakeholders from the community – as opposed to publishing a fixed, black-boxed system.

The way data is represented in Wikidata and Wikibase uses a flexible model of claims, references and qualifiers. Claims consist of items, properties and values, with properties acting as links that connect different concepts (items) to different values in the database, which could be other items, dates, URLs, plain text values and more. References add source information to claims, while qualifiers enrich claims with specifics such as the time period a particular claim can be considered valid, the specific type or role a value plays within the claim, etc. The structure of items, properties and values, as well as the terms used to denote these, can evolve and change to suit contextual requirements. Nothing has to be decided and fixed in advance by a central authority.

Still, decentralising the decision-making process as to how data is structured does not automatically guarantee a bias-free environment. In Wikidata’s open environment even non-logged in users can edit data and influence decisions about how data modelling is enacted in presumably democratic, but more often than not, culturally biased ways, too. The opacity of administrative procedures and the technical know-how needed to contribute to structural decisions in the platform tends to lead to the creation of homogenous and semi-closed, rather than truly open communities (Keyes, 2019).

Making LOD environments more reflective of the networked relations that are in place in a particular archive or collection of cultural heritage requires intentional actions on behalf of the curators, researchers and other institutional stakeholders that promote openness and community empowerment (Rossenova and Di Franco, 2022). Self-hosted Wikibase instances offer options for communities of art researchers and practitioners to self-organise around data modelling and data presentation while avoiding the pitfalls of the large public forum that is Wikidata (Fauconnier, at al. 2018). The following case studies illustrate three different approaches to working with difficult-to-describe networked objects and how they take advantage of and/or critique different aspects of the Wikibase and Wikidata software environments.

**Case study A: Rhizome’s ArtBase and the LOD approach to digital preservation**

Established in 1999, Rhizome’s archive of net art – the ArtBase – is an international and diverse archive (with over 2,200 artworks), primarily hosting works of net art, but also including works that employ materials such as software, code, moving images, games and browsers, among others. With its expansion in size and scope as well as complexity over the past 20+ years, the commitment to preservation of the works in the ArtBase has become an increasingly conscious effort at Rhizome.

Unlike plain text documents or static digital images, net artworks span beyond the boundaries of a single object and can become ‘diffuse’ (Fino-Radin, 2011), referencing external, dynamic and real-time data sources, or existing across multiple locations and platforms (Fig. 1). Capturing multiple layers of dependencies on media-specific properties of computers and networks, as well as various user interaction models, poses technical as well as social challenges when preserving
and presenting the net art archive online. Technical aspects related to the documentation in metadata and/or indication through interface design elements of unsupported browser plug-ins, changes in web languages or syntax, interaction models which develop over time and/or over various platforms, among others, were identified at the beginning of the ArtBase redesign project in 2016 (McKeehan, 2016).

The previous archival framework was no longer able to provide adequate access to a large number of the historic artworks in the ArtBase, and increasingly the focus of preservation efforts at Rhizome had been on building tools to support the presentation of complex artworks with multiple dependencies (McKeehan, 2016). Developments in browser-based emulation and web archiving tools had been instrumental in facilitating the restoration and re-performance of complex net art works as part of Rhizome’s major curatorial and preservation project launched in late 2016 – the Net Art Anthology. The remote browsing technology, first introduced in Rhizome’s oldweb.today project to emulate old browser environments, facilitated the online presentation of historic net art works in contemporaneous environments. Furthermore, the capacity to create high-fidelity archives of the dynamic web with Rhizome’s browser-based archiving tool, Conifer (previously Webrecorder), enabled the preservation of artworks based on third-party web services, such as Instagram and Yelp (McKeehan, 2016). These preservation efforts, however, had not been integrated with the ArtBase data structure and required a new approach to how contextual metadata around archives of web-based materials could be modelled and recorded in a collection management database.

In 2016, Rhizome was one of the first institutions to pilot the use of a stand-alone instance of Wikibase and to critically explore how practical LOD implementation might look like in the context of a digital cultural collection (Fauconnier, et al., 2018). Because every Wikibase installation requires a custom configuration of its data model, as part of the redesign project it was possible to develop a custom data model around the specific requirements of the net art works and Rhizome’s custom preservation and access policies. Moreover, this work was not done in isolation. It involved understanding and participation in the socio-technical processes of classification and maintenance that are entangled in the Wikibase project, its data architecture and adjacent communities, as well as other related LOD initiatives at the intersection of software and cultural preservation.
Many of the established metadata schemas or ontologies (i.e., the formal naming and definition of class categories and relations between data entities), from archival or museum institutions focus overwhelmingly on the description of physical objects and conform to a narrow conception of provenance, authorship and context around artworks that is not readily applicable to digital culture. Even models developed to take advantage of the affordances of LOD, such as CIDOC-CRM, focus predominantly on physical objects and the events that bind them in their physical environment (Linked Art, 2020). Models from the field of digital preservation and web standards have been focusing on the description of digital processes for some time already (PREMIS Editorial Committee, 2015; Moreau & Groth, 2013). PREMIS is an event-based standard for describing digital preservation metadata, which can be expressed as linked data and is widely used in the digital preservation community (PREMIS Editorial Committee, 2015). However, the level of event abstraction in PREMIS is of a degree more useful to managers of digital asset repositories than to scholars or curators and conservators who are concerned with humanities research, not just technical integrity (Rossenova, et al., 2019). Instead of adopting a ready-made ontology, the ArtBase data model work focused on selecting appropriate elements for description of net art works and their dependencies based on a combination of several heterogeneous models, including the Wikidata model for description of software objects developed by the Wikidata for Digital Preservation project (Thornton, et al., 2017) and the W3C PROV ontology for the purposes of expressing net art provenance – in particular, the relationships of derivation across multiple variants of the same artwork (Rossenova, et al., 2019), among others.

The new data model provides capabilities to describe the different instantiations of a single artwork in great detail. Each instantiation is accessible as a distinct ‘variant’ via new user interface templates (Fig. 2). These templates combine native Wikibase capabilities with additional functionality, needed to present complex assemblages of metadata, natural language text and images.
Alongside the Research & Development work done so far, since 2017 Rhizome has co-organized and participated in several dedicated events bringing together the Wikidata, Wikibase, and linked open data communities. In September 2018, Rhizome co-hosted the NYC Wikibase Summit at the New Museum, and led practical workshops dedicated to knowledge-sharing and prototyping solutions for the infrastructural, interface and data modelling needs of GLAM institutions. The results from these sessions were documented and further disseminated within the Wikimedia community contributing to subsequent policy discussions and strategies. Since 2020, Rhizome has been leading the development of the Wikibase Stakeholder Group dedicated to furthering the application of Wikibase solutions in the cultural and academic sector.

Although the ArtBase Wikibase remains closed for editing to external users and is mainly used as a presentation platform, Rhizome’s advocacy and development work across the Wikibase and Wikidata communities has led to significant changes not only in the codebase of these environments, but also in how communities, and more precisely GLAM institutions, interface with data in these environments. The LOD approach in this case showcases not only how one institution can model their data and/or interlink to external vocabularies and databases in order to preserve digital cultural heritage, but also how the challenges introduced by having to deal with this new type of networked materials can be addressed jointly by a diverse group of institutional and open source software stakeholders.

Case study B: The Digital Archive of Artists’ Publishing and the LOD approach to self-identification

The Digital Archive of Artists’ Publishing (DAAP) proposes an alternative model for the collection and description of artists’ publishing materials. The project was initiated by London-based, artist-led organization Banner Repeater around 2018 with a clear commitment to challenging the politics of traditional archives regarding inclusion and accessibility from a post-colonial, critical gender and LGBTQI perspective (Clarke, et al., 2021). Founded in 2010 by artist Ami Clarke, Banner Repeater is a contemporary art space featuring a reading room with a large archive of printed matter and an experimental project space in the midst of a busy public thoroughfare – Hackney Downs train station. Banner Repeater’s arts programme is driven by critical concerns regarding technology, and how historic experiments in text and publishing provide useful precedents for unpacking the contemporary socio-technical entanglements (Clarke, et al., 2021). The site of the project space provides open access to critical art and artist publishing projects for all, aiming to instigate new discussions and serendipitous encounters at a busy intersection of London’s public transport network. This approach to exhibiting a physical archive of publications led to the need for a similarly dynamic digital archive (of artists’ publishing) which could be co-created with artists and audiences in an open and collaborative process. This digital archive would include the materials already present at the physical site but would also be open to new contributions from local as well as international artist publishers.

Thus, even before the launch of the digital platform in 2020, the DAAP project deliberately gathered scholars and practitioners in the field of art, artists’ publishing and creative technology to contribute knowledge on the gaps in existing standards and knowledge management systems in order to set a precedent for modelling artists’ publications as linked open data (Rossenova and
Di Franco, 2022). This new model depends both on an alternative conceptualisation of the role and significance of publications within artists’ bodies of work, as well as on an alternative technical setup compared to typical library and museum cataloguing systems. Taking cue from Rhizome’s experiments with Wikibase as an open and flexible system, the DAAP also adopted Wikibase as an alternative to traditional CMSs, which could describe in detail the socio-technical context around artists’ published works, i.e., the multiple agents and components playing a role in the work’s production and distribution, while also complying with metadata protocols that make contextual description(s) accessible and discoverable.

Although artists publications do not face the same technical challenges as net art in terms of dependencies on software environments or third-party data sources, they still challenge traditional classifications and are not consistently catalogued by museums or libraries. The artist publication as an art object in its own right uses the discursive possibilities of printed (and sometimes digital) matter to open new avenues for collaboration and distribution that expand beyond the fixed form of a painting or a sculpture. Akin to installations, performances and even net art, artist publications can exist in multiple iterations spread over many locations and can involve many creative agents in their production (editors, designers, printers, other artists and/or writers). This propensity to ‘defy classification’ and to be collected and catalogued inconsistently by memory institutions poses challenges to how artist publications enter (or not) the art historical canon (Rossenova and Di Franco, 2022), but also opens opportunities to rethink whose work should be part of the archive in the first place and on whose terms. In order to customise the software and its internal data model to suit the specific needs of artists’ publishing, the DAAP involved not only technical development, but a wide range of activities focused on community building and collaboration around the design of the data architecture and, eventually, the access interface for users.

Following activities such as expert sessions with practitioners from the library, archive and publishing field, as well as public workshops and tutorials with artists, an initial beta version of the data model was defined. It specified some key items, properties and values, and recommendations on how to combine these into claims and eventually statements. Still, the structure of properties and values within these statement categories need not remain fixed. As the material components of the works change (e.g., increasingly involving digital files, alongside finished printed products), or new communities emerge around the publishing process (e.g., communities from diverse and multilingual backgrounds who have their own vocabularies when it comes to the description of both publications and publishers alike), the schema of the archive database can evolve accordingly: new items can be added as well as new properties which link items; or the meaning of existing properties can be adapted to new contexts. The collaborative editing and version control features of Wikibase mean that, as the data model evolves, earlier versions are not lost and do not become obsolete.

The open sessions involved individuals interested in contributing their own works to the archive and facilitated open conversations with a diverse range of participants around the choice of specific categories for describing works, the methods for attaching contextual narratives to publications and, last but not least, the classification of artists, collaborators and publishers present in the database. The participatory methodology aimed for the inclusion of multiple voices and concerns and for the possibility for artists traditionally excluded from the canon to self-identify their works and themselves. They could create new items in the Wikibase database that
describe their artworks and themselves (avoiding the pitfalls of bias encountered in public platforms like Wikibase and Wikidata). Facilitating this process involved not only adapting the data architecture of Wikibase, but building an additional frontend interface (Fig. 3) and developing in-depth training materials that would make the LOD environment more accessible to users of the archive who are coming from visual arts backgrounds as opposed to computer or information science.

Figure 3. Illustration of the DAAP interfaces – Wikibase (left) and custom web frontend (right). The magenta lines trace data points pulled from the Wikibase database into the web frontend.

In developing both an accessible LOD environment and custom user workflows around it, the DAAP project aims to make archival labour more open, collaborative and non-hierarchical and to provide new possibilities for artistic encounters and for self-identification of artists with artists’ publishing materials.

**Case study C: The Joan Jonas Knowledge Base and the LOD approach for data reuse**

The third case study focuses less on building a custom data model in a private Wikibase instance and more on exploring the possibilities to make research data about complex artwork objects more accessible and easily reusable by external audiences. For this purpose, the project explored the potential role of Wikidata not only as a repository to offload digitised collections data (as has already been the case with various museum collections), but also as a long-term, supplementary resource for research which can be utilised in critical ways. This case study is the LOD component of the [Joan Jonas Knowledge Base](https://joanjonasknowledgebase.org) (JJKB), a resource built by the [Artist Archives Initiative](https://artistarchivesinitiative.org) based on their approach of focusing equally on humanities-based research and implementing a complementary underlying digital infrastructure (Artist Archives Initiative, 2023). JJKB houses information about the New York–based artist Joan Jonas (b. 1936) who has been at the vanguard of interdisciplinary art forms such as performance, video art and new media installations for over five decades (Rossenova, et al., 2021).
The project involved close collaboration between the curatorial and conservation teams and the technology team in selecting and preparing a unique research resource drawing on materials from the artist’s personal archive, as well as museum archives, archives of photographers, galleries, university libraries and other public and private archives and foundations holding relevant documentation of the case studies, photographs, videos, publications and exhibition ephemera included in the JJKB. Throughout this process it became clear that performance artworks are associated with many entities, including objects such as drawings, audio, video and other media; iterations of performance over time; and collaborators including artists and institutions. The need to capture and then be able to access and work with this complex network of iterations, related objects and collaborators requires an alternative to a traditional database solution.

At the same time, in accordance with the core vision of the Artist Archives Initiative, the project needed to rely on software tools that could be sustainably maintained over time, and serve as a model for other arts, cultural and research organisations. Wikidata was chosen as the solution for the LOD component of the project since it could store data over extended periods of time independent from research project–funding cycles; Wikidata’s open SPARQL endpoint could then be used to query the data; and a WordPress front-end completed the project’s technology stack in providing users with a more familiar and accessible web interface. Research data that was originally compiled by the curatorial team was discussed collaboratively with the technology team in order to coordinate what data was going to be used for the final dataset to be uploaded to Wikidata. That dataset was then further enriched with additional data that was already in Wikidata, such as bibliographical data about people, artwork or exhibition data from some of the institutions that had collected or exhibited Joan Jonas’ works, geo data about locations as well as external identifiers linking to various authority control data sources.

To be able to integrate the JJKB dataset into Wikidata’s existing data model consisting of properties and items, the LOD team had to work within the constraints of the system and test its limits when it came to description of non-traditional art objects. After consulting available reference materials for community data curation within Wikidata (e.g., the WikiProject page for Visual Arts, or the Rhizome-initiated WikiProject for Digital and Performative Arts) a limited set of properties and item classes were selected and specific ways to relate these properties to data in the JJKB dataset were defined (occasionally going against the grain of how these might be used within the Wikidata ecosystem as a whole). The aim was to align and map the use of properties as much as possible to data regarding Joan Jonas and her works already entered within Wikidata (e.g., by curators from the Tate or MoMA whose collections data was already in Wikidata), in order to make connecting and cross-referencing with data from other institutions more accessible.

An important question from the start was: How to bind a small dataset amidst the vastness of Wikidata? When reusing the data later to search or visualize it in specific ways, it would be good to be able to request all data from the JJKB without missing something. Equally, on occasion it might be helpful to distinguish between data that comes from the research team of the JJKB vs data that comes from other sources. The solution developed in consultation with other institutional representatives was to add the statement “described by source: JJKB” to all new items created for the JJKB research dataset, although this was not ideal in some cases (e.g., for
artists and exhibitions that span beyond the boundary of the JJKB, such as Richard Serra or documenta 6, whose item pages contain a lot of data outside the scope of the JJKB).

A further concern was working within data-modelling constraints set by other knowledge domains which do not meet the needs of performance and avant-garde art practices per se. For example, the constraint that the location of an artwork cannot be different from the location of the collecting institution does not always work for performance artworks, which may enter various collections after the performance itself took place. In some cases it was possible to use workarounds, but workarounds limit the possibility to fully query the data and it would be good to discuss further with the Wikidata community longer-term solutions, such as changes in the core data model.

Exploring the limitations of the current data model in Wikidata produced useful insights regarding how other art organisations could also use the public knowledge base, which were documented in detail on the JJKB website. But beyond focusing only on the limitations, this process also highlighted how interlinking art research data amidst the network of many other knowledge domains, collections and institutional datasets can take advantages of the reusability of LOD in ways that support further research. This is made evident when accessing Wikidata’s SPARQL endpoint to ask questions of the research dataset and to subsequently render data visualisations (Fig 4).

Figure 4. Illustration of the DAAP interfaces – Wildbase (left) and custom web frontend (right). The magenta lines trace data points pulled from the Wildbase database into the web frontend.

All queries and corresponding visualisations for the JJKB were developed through a highly collaborative process across all teams on the project. It was important to design the queries and the default visualisations chosen for each query with the needs of researchers in mind, while maximising the potential of each query, and each visualisation, to showcase the varied, complex and at times surprising connections that could be discovered across the JJKB dataset and related data in Wikidata. The project’s technical director, Deena Engel, also developed a python script that pulls the SPARQL query code to generate HTML pages on the JJKB website for each query; this keeps even the static frontend presentation ‘lively’, i.e., pages can be automatically re-generated if there are new research updates.
The LOD work in the JJKB project illustrates that, to be able to work effectively with LOD in an open environment like Wikidata, it is very important to prepare and model data in collaborative ways and with a view to interlinking and reusing as much of the existing infrastructure as possible. Beyond questions of technical infrastructure and classification standards, the project had to grapple with ethical considerations when it came to how data is shared and represented in Wikidata, as well as sustainability considerations and making one’s data useful and usable to a variety of interested communities. All project findings and recommendations in the forms of toolkits and tutorials aimed at others studying contemporary performance art were made available via detailed documentation on the project’s website (The Joan Jonas Knowledge Base, 2023).

**Summary and recommendations**

This essay showcases parallels between the tasks of archiving net art, artists publications and research data related to performance art and the task of archiving artists’ websites by drawing upon findings from three distinct case studies. What the case studies all have in common is the critical engagement with Linked Open Data (LOD) software environments as a means to structure, classify and store heterogeneous data. Artists’ websites as sites of art and digital cultural memory located at the intersection of various material and immaterial networks of relations (e.g., web protocols, software, artist collaborations, institutional events and more) bear much in common with the art objects, and subjects, present in the case studies of Rhizome’s ArtBase, the Digital Archive of Artists’ Publishing and the Joan Jonas Knowledge Base. Across all these cases, networks or relations play significant roles in how artworks are produced, distributed and subsequently archived, preserved and made available to the public again as reperformances or datasets for research. LOD, which relies on an equally networked structure of an underlying graph database infrastructure in order to make possible the interlinking of entities in flexible, non-hierarchical ways, is well suited to meeting the cataloguing needs of these networks of relations. The open source software project Wikibase, and its public instantiation Wikidata, provide a practical and sustainable LOD implementation for the broader GLAM and academic research sector.

Although the case studies in the essay focus primarily on the question of how LOD can address challenges associated with the cataloguing of diverse art objects and sites of art, the characteristics of assemblages and networks exhibited in these challenges also lead for further considerations regarding the ethical, technical and curatorial choices. The case studies also raise relevant questions concerning long-term maintenance of archival projects that fall outside the remit of traditional memory institutions, as well as the need to adjust institutional policies in response to the changing contexts of art objects that are no longer produced in an isolated manner using the familiar mediums of, e.g., the painting, the sculpture or the printed monograph. The aim of the essay, however, is not only to highlight the problems of attempting to archive non-traditional art forms, such as artists’ websites, but also to open space to discuss the opportunities that emerge when research projects engage with open source software and related communities in the case of Wikibase and Wikidata. Working with open, collaborative systems offers new opportunities to rethink communications between art and archive caretakers and the artistic community whose work is being archived and preserved, by asking: How could community voices be made more visible in the archive infrastructure and interface? Last but not least, the essay
hopes to inspire change in institutional policies related to the status of materials that may be considered ‘secondary’ – websites, documentation, publications and other ephemera. If the latter are considered important sites of art in their own right, the valuable relations that can be captured and described in the archival records for these materials can contribute to preserving a rich context around artwork production and artist's communities spanning across institutional boundaries. Reuse of such contextual data via LOD environments opens up new possibilities for research into contemporary modes of art making that are distinctly networked and not bound to the singularity of traditional object representation.

**Bibliography**


