(Re)capturing the Emotional Geography of Lost Music Venues: A Case Study of the Willow Community Digital Archive

Rachel Cowgill
rachel.cowgill@york.ac.uk
School of Arts and Creative
Technologies,
University of York, UK

David Bainbridge
davidb@waikato.ac.nz
School of Computing and
Mathematical Sciences,
University of Waikato, NZ

Alan J. Dix
alan@hcibook.com
Cardiff Metropolitan University,
Wales, UK

Victoria Hoyle
victoria.hoyle@york.ac.uk
Department of History,
University of York, UK

Vicki Fong
vicki@thomfong.com
ThomFong,
Southwark, London, UK

David Thomas
davied@thomfong.com
ThomFong,
Southwark, London, UK

ABSTRACT

The loss of many high-street music venues in recent years has highlighted their connectedness to place and communities. Understanding the emotional geographies of these venues, as experienced by their patrons, is key to explaining the outcry that can accompany such closures. In these circumstances it can be challenging to try to (re)capture the intangible elements that defined a lost venue and widen the scope for musicological enquiry. This paper sets out to address that challenge by exploring methods developed by the Willow Community Digital Archive to co-create a community archive in celebration of The Willow, a family-run restaurant-cum-nightclub that operated in York, UK, for over 40 years. Further, we report on how these methods informed the crafting of a general-purpose digital library system to form the archive. We also detail some initial experiments with ChatGPT, embedded into the archive, to investigate its potential to encourage visitors to engage with and inspire further contributions to the archive.

CCS CONCEPTS

• **Information systems** → **Digital libraries and archives**; • **Applied computing** → Arts and humanities.

KEYWORDS

Music Venues, Nightclubs, Heritage, Data-gathering, Digital Archives, ChatGPT, DJ-ing, British Chinese Communities, York

1 CONTEXT

The past decade has seen a steepening decline in the fortunes of music venues across the UK. For 2023 alone, the Music Venues Alliance reported a drop in the membership of the Music Venues Trust of 13%, from 960 to 835, with 51% of those 125 venues having ceased trading and 49% no longer hosting live music [4, pp. 9, 14–15]. As we go to press, the House of Commons Culture, Media and Sport Select Committee has published its report on the escalating crisis [14]. A gruelling combination of economic and social impacts arising from pandemic restrictions, increasing rents and energy costs, noise-abatement complaints, planning disputes, and evictions for redevelopment has been taking a heavy toll on promoters, artists and communities who build their lives and livelihoods around venues, as well as the ecosystems they provide for the support and development of new creative talent. With closures comes loss—of physical spaces and material traces, but also of the particular mix of sounds, smells, tastes, looks, moves, interactions, memories, sensations and emotions, which when experienced cumulatively and collectively creates something perceived as special to a venue. For a lost venue the latter is particularly hard to (re)capture, given the original space may be demolished, out of bounds or changed beyond recognition; but doing so is crucial to understanding the significance of a venue in a particular time and place, the emotional attachments it can inspire, and strength of feeling that can accompany its demise.

Addressing this challenge became a priority for the project we report on here, which set out to create a community digital archive of The Willow restaurant-cum-nightclub in Coney Street, York (UK). For many The Willow was a cherished feature of York's city-centre nightlife, achieving remarkable longevity of more than 40 years in a sector not known for its stability, and the shockwaves caused by its sudden closure in July 2015 were felt as far as the national press [17]. The project demonstrates a 'concern with the spatiality and temporality of emotions, with the way they coalesce around and within certain places', that Bondi *et al.* associate with the recent 'emotional turn' in urban and social geography—specifically the emergence of 'emotional geography', characterised by 'a non-objectifying view of emotions as relational flows, fluxes or currents, in-between people and places rather than "things" or "objects" to be studied or measured' [11, p. 3]. Research into people's experiences of and affective responses to night-time leisure spaces has tended to rely on surveys and recorded, transcribed, tagged and anonymised

Figure 1: Photograph of 37a Coney Street, York, where The Willow was located before its closure in 2015.

interview data (as in, for example, [19]); however, communities that coalesced around a venue that is now lost, such as The Willow, can no longer be accessed via a physical and temporal point of connection. Other approaches to gathering data therefore needed to be found.

In the first part of this paper, we explore the complexities involved in researching and representing lost and historical music venues, and propose methods and techniques suitable to be deployed, combined and/or modified for this purpose: a triangulation of physical traces, traditional archival sources and personal memories accumulated and co-produced through a variety of physical and digital interactions between community and university researchers—all facilitated by a pop-up exhibition space on the one hand and a dynamic, purpose-built digital archive on the other. In the second part we explore the structural and technical development of the Willow Community Digital Archive (WCDA) itself, which centres on the configuration and deployment of a general-purpose digital library software system. From there we took things a step further, with some preliminary experiments where we embedded a custom ChatGPT-base assistant into the digital library, with a view towards using it to engage visitors in narrative-building around content and invite new affective responses. This WCDA project expands on work the authors have produced in relation to the InterMusE project (such as [5, 6, 15]) framed within recent popular-musicological interventions in critical heritage and archive studies [9, 12, 13, 22].

2   LIMINALITIES

Data gathered via the WCDA project over a period of eighteen months enable us to provide an introduction to The Willow—a history which, in turn, sheds light on the experiences of three generations of the British Chinese community in York. Co-proprietor of The Willow, Tommy Fong, is the grandson of a Chinese seaman who travelled to and from the UK for work in the 1920s. His father was brought to the UK to study, returned to China to marry, and then settled with his wife in Hull in 1936. From there the family moved to Liverpool, Middlesbrough and Darlington, where Tommy was born in 1945 and named after the British soldiers who were

arriving home from war. Tommy worked from childhood in the family laundry and wooden-toy-making business, and initially trained to be an accountant; but in 1973, he opened a Cantonese restaurant in York's historic city centre with his brother and brother-in-law,[1] bucking the then trend among Chinese migrants for opening 'take-aways' [10], and employing six chefs to deliver 'delicacies similar to those served to the Emperors of ancient China [...] handed down from generation to generation for over a thousand years' [1]. Such rhetoric not only stressed authenticity, but also resonated with the city's Roman and other ancient heritage in ways the Fong brothers probably hoped would appeal to prospective customers. The premises were upstairs at 37a Coney Street (see Figure 1), formerly the well-appointed Willow Café, established in 1936—some of the café's original chairs, silverware and willow-pattern crockery were retained along with its name. English dishes were also available for diners, as was advice and reassurance on how to eat 'communally' for those unaccustomed to the Chinese way of sharing food [1]. At The Willow Tommy met his future wife, Soo Mei, who would join him in the business: she waited tables there after leaving her rural village in Hong Kong for the UK in 1973 and lodging with relatives in York [18].

Music was a key component of The Willow from the beginning: inspired by watching Fred Astaire and Ginger Rogers movies, Tommy's adverts invited customers to 'revel in the night-club atmosphere' and in 1974 to 'Dance in Candlelight into the early morning with our resident band', a quartet named the Alan Scot Sound—'(real funky!)' [2]. Chairs and tables were positioned around the edges of the dance floor. Live music proved difficult to manage, however, and was replaced by DJs from Hospital Radio who donated their fees to charity. When competition from York's first 'all-you-can-eat' Chinese buffets forced the restaurant side of the business to close in 2008, The Willow began to operate solely as a night-club/disco, albeit with a steady supply of complementary prawn crackers [3, 18]. Increasingly it catered for students (especially student societies) and workers in the hospitality and entertainment sectors who needed a place to socialise after finishing work. Strong relationships developed over the years with the city's two universities and York Theatre Royal, as new students were brought to The Willow in Freshers' Week by peer-mentors, and YTR cast and drama/panto-society members performed routines and set pieces on The Willow dance floor, and referenced The Willow in their shows. The DJs catered for wide-ranging musical tastes within the broad category of commercially produced pop, and based their sets mostly on customer requests, creating the impression of free choice—'anything could get played' [3]; but Soo Mei recalls her husband's strictness with the DJs—he would have a word if they played music he did not like, and if they did so for a second or third time, 'he would scream' [18]. "Out of Space" (1992) and "Firestarter" (1996) by the British band The Prodigy were among the tracks disallowed in this way. DJ Max would also include Tommy's personal favourites, such as songs by The Beatles (Tommy was a 'massive fan') and Carl Douglas's "Kung Fu Fighting" (1974), a global-music mashup whose opening flute solo outlines a stereotypical 'oriental'

---

[1]Tommy's brother and brother-in-law soon left the business, but as his name was on the lease he continued. His brother became an academic at Durham University, specialising in theoretical physics and cosmology, having studied at Cambridge and the University of Maryland—the late Dr Richard Fong (1936–2017).

signifier (in Western musical terms). Tommy's 'obsession' with this song was also celebrated in a charity Karate Night featuring martial arts displays at The Willow in 1984 [3, 18]. Tommy's monitoring of the soundtrack aligns with his broader approach to venue management—his familiar presence, constant vigilance, often controversial dress-code and door policies ('students and regulars') and quiet removal of customers exhibiting drunken behaviour are credited, particularly by the club's LGBTQ customers, with creating the atmosphere of freedom and safety associated by many with a night at The Willow.

Six years after the closure of The Willow, in 2021, an interdisciplinary team of researchers from the University of York, working with the City of York Council and supported by a grant from the UK CRF scheme,[2] opened a pop-up venue in an empty retail unit just three doors down from the site of The Willow. The project—StreetLife: Using York's Historic High Streets as Heritage Catalysts for Community Renewal—promised to 'create innovative, immersive experiences to revitalise Coney Street, combining digital innovation and physical engagement', and to 'forge links between retail premises and creative, heritage-led renewal'. Coney Street is York's most prestigious and intensely commercial thoroughfare, with one of its highest footfalls and a history stretching back over two thousand years; but like many UK high streets it has lost much of its vitality through steady decline.

By rediscovery of its forgotten histories and communities, StreetLife undertook to change people's perceptions of Coney Street and imagine a new future for it inspired by its past. The empty shop (the Hub) offered a blank canvas—a neutral, liminal space for co-creating exhibitions, workshops, performances and other activities shaped by visitors' memories and conversations about the street. By far the highest proportion of musical memories, and eliciting the most emotional responses, were those concerning The Willow and its 'legendary' status in York's collective imagination.

3 RECONNECTIONS AND RECOLLECTIONS

The challenge with investigating any lost venue is how to connect with past patrons, promoters and performers when the venue itself no longer draws them together in real space and time. This problem was particularly acute for the StreetLife project, whose initial funding from the UK CRF fund was for six months only. A small team of researchers and student interns was recruited, including several who had attended The Willow, but also several who had not, and so could question respondents without preconceptions. Our priority was building-up an iterative cycle of engagement, both physical and virtual, inspiring contributions of new layers of material in response to what had already been collected, and creating a series of 'entry points' for those drawn into the project later by friends and family (snowballing). For a chronological account of how this process unfolded, see Table 1.

Different ways of contributing were offered depending on preference and opportunity. This included focus groups, semi-structured individual interviews, meet-ups in the Hub, and walking trails. New conversations could be opened informally, and photographs played a prominent part in communications, both as artefacts in themselves

and as a means of digitising artefacts. We had a breakthrough moment when Vicki Fong agreed to mediate between the project team and her family: her parents, in their sixties and seventies, were supportive of the project but did not initially want to be involved; now, however, we have extensive oral-history materials via their daughter as interviewer. She has also shared her own original artwork from The Willow, gathered artefacts, and with David Thomas (as ThomFong) created content to encourage new thematic strands to emerge.

Our methods also brought together a rich archive of physical and symbolic artefacts, for example the Willow t-shirts (designed by Vicki), pass-out stamps in green ink (large and multiple—still very visible the day after), and Willow street signs, in addition to memories. The next step was to find a way of bringing all of this diverse material into dialogue, and by so doing to reveal an 'affective community' focused on The Willow—what Zink describes as 'the temporal, spontaneous, and generative solidarities that bind affected and affecting bodies and which give rise to intense and immersive forms of integration and belonging' [24].

In pursuit of this goal, we surveyed general-purpose Galleries, Libraries, Archives and Museums (GLAM) sector software [21]. Through this the opportunity arose to work directly with David Bainbridge, one of the key developers of Greenstone, a prominent Open-Source Digital Library Software toolkit [23] to create a digital archive for The Willow.

4 BUILDING THE ARCHIVE

Greenstone has a pedigree of over two decades of R&D. On the production-ready 'development side', the software interface has been translated into over 60 languages, and has a track record of being used by organisations both small and large: from the University of the Antilles and Guyane's digital library *Manioc: Bibliothèque numérique, Amazonie, plateau des Guyanes* (Digital library on the Caribbean, the Amazon, the Guyana Plateau),[3] for instance, to national-institution-scale digital libraries, such as Singapore National Library's *NewspapersSG*.[4] The toolkit is agnostic as to the metadata standards used, including built-in support for Dublin Core,[5] and MARCXML,[6] for instance, and export capabilities including METS.[7] Content can also be served via a built-in OAI-PMH[8] server, if desired, to aid interoperability.

Pertinent to the Willow Community Digital Archive project, Greenstone's off-the-shelf capabilities include the ability to process all the amassed files assembled by the project: Excel spreadsheets; Microsoft Word and Adobe PDF documents; and the multimedia formats PNG, JPEG, MP3 audio, and MPEG video. Drawing upon a suite of open-source command-line utilities that can process a range of file formats, Greenstone's ingest process provides a flexible workflow for how these file formats are processed and combined: text content from the Microsoft Word and Adobe PDF documents can be extracted and used for full-text searching in the digital library; multimedia content can be processed by Greenstone to produce

---

[2]https://www.gov.uk/government/publications/uk-community-renewal-fund-prospectus

[3]http://www.manioc.org/

[4]https://eresources.nlb.gov.sg/newspapers/

[5]https://www.dublincore.org/resources/metadata-basics/

[6]https://www.loc.gov/standards/marcxml/

[7]https://www.loc.gov/standards/mets/

[8]https://www.openarchives.org/pmh/

May 2022: Filming of 360-degree footage by StreetLife inside the former Willow premises at 37a Coney Street, midway through conversion into a café, courtesy of current proprietors Coopland & Son (Scarborough) Ltd;[a] edited into VR and played on headsets in the StreetLife Hub (credits: Lucy Barker (film), Adam Gill (VR)).[b]

June–November 2022: Desktop and archival research by StreetLife researchers and UoY (University of York) interns: press and web coverage, social media, image collections, student newspapers (Borthwick Institute of Archives and York Explore), Fabrication Crafts' Coney Street Heritage Project website,[c] and playlists in Spotify[d] and other formats.

June–November 2022: First wave of interviews, with DJ Max and patrons—snowballing, with interviewees drawing on and sharing personal photo collections—by StreetLife researchers and UoY student interns.

July 2022: Willow Timeline produced on Padlet and included as an example in the Our Place online DIY community digital-archiving toolkit for music venues (credits: Charlotte Armstrong, Lizzie Hodgson, Andy Egerton, Frankie Perry).[b]

July 2022: Willow community meet-ups[e] at the StreetLife Hub with the Willow Timeline reproduced on paper and mounted on the walls, with empty sheets, pens and post-it notes for visitors to the Hub to leave additions/annotations/memories; early trials of TalkOver[f] application for linking oral history and photographs [15]; Google Form launched for gathering memories remotely and promoted in a blog post by StreetLife researcher Lizzie Hodgson, 'The Willow: Gone but not Forgotten'.[g] Also featured Willow playlists and complimentary prawn crackers.

July 2022 and ongoing: Initial contact with Vicki Fong, daughter of the proprietors of The Willow, Tommy and Soo Mei Fong, via York Music Venue Network (YMVN), leading to her interviewing family members, contributing many artefacts and personal recollections, and later collaborating in design and promotion of the WCDA as a partner in the strategic-design duo ThomFong.[h]

August & November 2022: Focus groups and semi-structured individual interviews with York Theatre Royal staff around memories of The Willow (by StreetLife researchers and UoY student interns).

September 2022: Publication of *Coney Street Highlights* and *Echoes of Coney Street*, self-guided multimedia walking tours on the StreetLife website, including The Willow as a waypoint.[i]

November–December 2022: Publication of *York's Historic Music Venues Trail*, a self-guided multimedia walking tour on the StreetLife website, including The Willow, co-curated by Lizzie Hodgson and Chris Sherrington (YMVN).[b, i]

January–March 2023: *A Queer Walk of York*, guided and self-guided tours of York's historic LGBTQI+ sites, including The Willow, published on the StreetLife website; co-curated by StreetLife researcher Andy Egerton ('live' guided tours led in June 2023 and June 2024).[i]

January–June 2023: *Love it or Hate it? The Willow Legacy Art Exhibition* (StreetLife Hub) curated by Andy Egerton, featuring banners, t-shirts and artwork created for The Willow (including the famous 'Love it or Hate it' image) by Vicki Fong and photography by Ceri Oakes, and others.[j]

February–July 2023: Design and creation of the WCDA, through collaboration between StreetLife, the DL Software Development Team, ThomFong, York Explore, and the UoY Institute of Public Understanding of the Past.

March–April 2023: *Archive All Areas: People, Places, Memories and Music*—a DIY exhibition celebrating York's music venues, promoters, and musical ephemera, co-curated with YMVN at the StreetLife Hub (credits: Chris Sherrington, Jenn Chubb, Suzy Harrison, Liam Maloney, Yorgos Pachos, Andy Egerton).[k]

May–June 2023: Creation of six curated video stories by ThomFong, drawing on artefacts from the WCDA to promote engagement with the WCDA for use on social media.

2024–25: publication and promotion of the WCDA via the StreetLife website with added mechanisms for gathering further comments and feedback, and promotion of Our Place, the DIY community digital-archiving toolkit, among music-venue owners, promoters and patrons.

2026: Planned transfer of the WCDA to York Explore Libraries and Archives for preservation in 2026 (using Preservica).

[a] https://planningaccess.york.gov.uk/online-applications/ (see planning case 17/02976/LBC)

[b] https://ourplace.whereweare.org/examples/

[c] https://coneystreetheritageproject.org.uk/

[d] For example, 'Willow Disco York – Love it or Hate it?', by *parmahamsam*, https://open.spotify.com/playlist/1LTkdULNt0bShZvLWfl2rK

[e] https://www.streetlifeyork.uk/events/willow-community-project-meetup

[f] https://alandix.com/labs/talkover/

[g] https://www.streetlifeyork.uk/discover/willow-gone-not-forgotten

[h] https://www.thomfong.com/

[i] https://www.streetlifeyork.uk/explore

[j] https://www.streetlifeyork.uk/events/love-it-or-hate-it-willow-legacy

[k] https://www.streetlifeyork.uk/events/archive (see 'Archive All Areas')

Table 1: Willow Community Digital Archive: iterative cycle of data and artefact gathering and presentation (physical and digital) since May 2022—via the StreetLife Hub, social-media accounts and website. Credits here (as throughout this paper) are in addition to contributions from the co-authors.

derivatives more suitable for consumption via the web; and the data stored in spreadsheets can be made available as catalogue records of metadata in the digital library, which can subsequently be used in browsing, searching, and faceting features. This ingest workflow is controlled through a Graphical User Interface (GUI), designed for GLAM users. The GUI also allows the customisation of searching, browsing features, and the presentation of documents.

Over on the 'research side' of R&D for Greenstone, the core software architecture—which covers the ability to expand importing and exporting documents through Plugins and Plugouts respectively—already provides a versatile environment in which to go beyond the status quo. Added to this, the software provides an extension mechanism which allows for the augmentation of, and changes to, build-time and runtime functionality. This allows for the exploration of research ideas that go well beyond the software's normal mode of operation, and can be done in a way that does not interfere with the standard deployment of the software [7, 8].

Chronologically, in the development timeline of the Willow Community Digital Archive project, having established the core online digital archive using standard features of Greenstone (which we detail in Section 4.1, below), we then leaned into the research side of the project, and experimented with integrating a ChatBot capability into the archive, using the Assistant API, available in beta-form from OpenAI. In Section 4.2, we show how we integrated this chat capability with the Greenstone software architecture. We also provide some observations on the interaction experience users have with 'Willow Sage', the name we gave to the developed Greenstone DL collection + ChatBot capability.

4.1 Using a general-purpose digital library toolkit

As previously mentioned, in terms of ingesting all the files that had been amassed for the Willow project on GoogleDrive, all file formats were supported by Greenstone. A custom Greenstone *interface* was created for the Willow Community Digital Archive project, along with a custom *site* within which a collection called *community-contribution* was formed.

Keeping within the standard 'off the shelf' capabilities of Greenstone, a document-processing plugin pipeline for *community-contribution* was created to ingest the collated shared-drive of content. Options provided to plugins as part of their configuration allow their behaviour to be fine-tuned. Notable application of these options in the pipeline include:

- Adding a *tw.* prefix to the metadata extracted from the spreadsheet files to form a namespace for *the willow* metadata set;
- Binding spreadsheet metadata to a document based on the *Filename* column heading;
- Creating a surrogate document out of the metadata fields in a row of spreadsheet data if it does not match the *Filename* column heading (or else the cell is empty);
- Overriding the 'hash on file content' default behaviour used by Greenstone to instead set the ID to be that provided by the metadata value *tw.DLIdentifier*, resulting in more readily understood document URLs produced by the DL software; and

- Overriding the default size of thumbnail derivatives generated by *ImagePlugin* to be larger and more inline with the resolutions of screen used by the target users for the project (moving away from the more conservative default value which is aligned with the software's use by UN agencies on legacy PCs in developing countries).

In terms of displaying pages within the Digital Library, Greenstone uses eXtensible Stylesheet Language Transformations (XSLT) [20] to control the appearance of pages in the digital library in a customisable way. The runtime system generates XML-based messages, which are converted to HTML through the application of XSLT files. Core XSL files for a collection in the *default* interface covers browsing, search, document presentation, and an 'About this Collection' home page. For core XSL files, any of the *<xsl:template />* rules in the given XSL files can be overridden in a custom interface, such as the one we developed for the WCDA. Additionally, if a collection specifies any templates in an XSL file, these are used in preference to the ones in the interface. There is also provision for adding in XSL files for wholly new pages.

All of these capabilities were used in forming the WCDA interface. The main customisation work occurred in the *web/interfaces/thewillow* area of the DL's filesystem, the majority of which was realised through the application of CSS files. The desired level of control of the presentation of documents was significant enough to warrant a custom interface *document.xsl* file. Similarly, the required search functionality for the archive was significantly different to that provided by the default interface, which was also met with the addition of XSL files. We discuss these two examples below.

Even though the DL consists of only one collection—and so in theory all decisions related to presentation could have been handled by the custom interface—in situations where we saw the requirement to be collection-specific in nature, with the aim of future-proofing, we codified that *<xsl-template />* in the DL collection's *community-contributions* area of the file-system.

Figures 2 and 3 show snapshots taken from the developed DL. The collection home page to the Willow Community Digital Archive (not shown) provides background information about the project, and includes an introductory video. Through tabs in the header banner of the archive, a user can go beyond the home page and access: Stories; Explore More; Add Your Memories; and Chat with Willow Sage.

The first of these (Stories) is an example of a browsing feature, built around metadata provided when the collection is ingested. Figure 2 captures the moment where the user has clicked on the tab for the first time: it loads a new page in the browser, which alphabetically lists all the videos that have been added to the collection, along with a popup window that appears over the *Stories* tab as a quick-access feature to access a specific video, without the need to scroll down to find it.

Through the *Explore More* tab, a user can search the archive. Figure 3 shows an example where the user has entered the term 'max' into the search box (the name of the main DJ who worked at The Willow in later years) and ticked three (of the six possible) Category check-boxes: Interviews, Photographs, and Memories. Hierarchically organised by folder on the shared drive, 'Interviews'
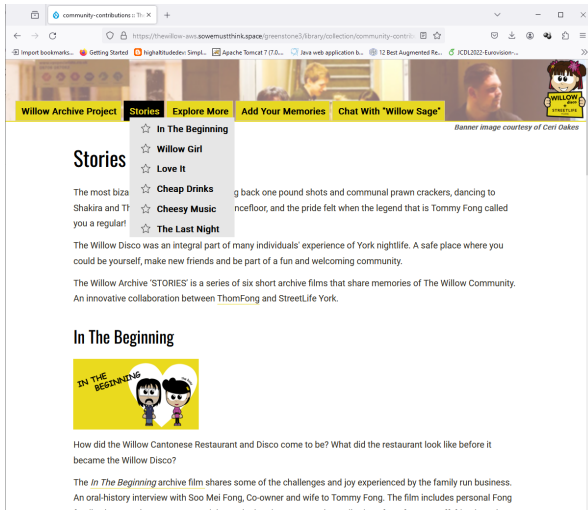
Figure 2: Browsing the WCDA by curated video stories.

Figure 3: Searching the digital archives for references to 'max' (to find material related to 'DJ Max'), restricted to the document categories Interview, Memories, or Photographs.

is metadata assigned to the Microsoft Word documents that resulted when researchers on the team conducted semi-structured interviews with previous staff and patrons of The Willow who had responded to a request to come forward and help contribute their recollections (see Table 1). 'Photographs' are images (again marked out through metadata) that are the result of the endeavours of researchers in curating the archive, or else photographic content that was brought along by former staff and patrons as part of the interview process, which was subsequently digitised. The project has also developed a Google Form used for soliciting (asynchronously) recollections by former staff and patrons. The document workflow used here is one where a newly submitted form is first QA-checked by one of the curators of the WCDA, and if all is as it should be, then the submitted document is added to the Google Drive, and metadata related to the submission added to the relevant spreadsheet (also stored on the shared drive). Next, utilising Greenstone's ingest pipeline, the WCDA is rebuilt, with the new submission(s) added to the archive, along with metadata assigned so they are searched when the checkbox 'Memories' is selected.

When the 'Search' button is pressed in Figure 3, behind the scenes—using a *<xsl:template/>* in the collection's *query.xsl* file that overrides Greenstone's default behaviour—the text in the search box and the status of the checkboxes are looked up, from which an advanced query syntax string is produced, which is then used as the query submitted to the digital-library server as part of its search action.

At the direction of the design-team members, additional XSL templates were overridden in the interface's *query.xsl* file so that results were displayed in a tabular layout below the search-box area (rather than the Greenstone default, which displays the result-set vertically) and makes prominent use of textual metadata. For the WCDA, large visual icons were included in the artwork supplied for representing categories such as the word-processor documents, matched to be the same size as the thumbnail images. The result set displayed was further customised to use these images as the key element displayed in the tabular layout, with 'alt text' attributes to each
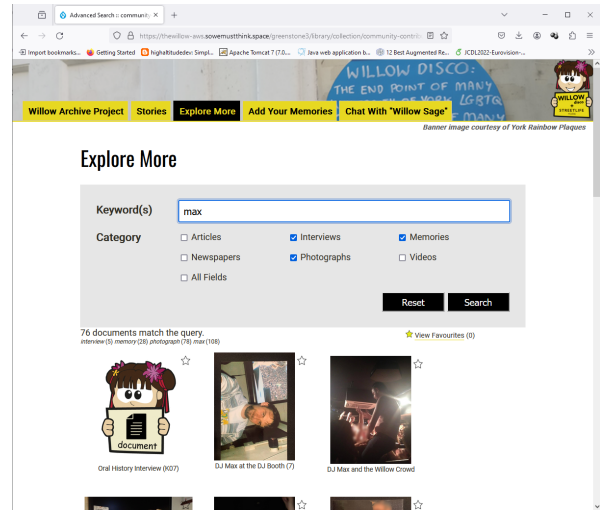
image set to title metadata, so that, when a user hovers over an item, a tooltip appears providing text that describes the item. Putting this all together, for the entered search term 'max' in Figure 3 with the given Category settings, 76 documents were found. The first matching item is one of the formal interviews (Oral History K07), with the following items that are visible in the snapshot being photos.

4.2 Experiments with an embedded Chatbot

In November 2023, OpenAI released the Assistants API for their platform,[9] through which you can develop your own custom GPT-powered interaction experience for your users. Producing a specialist chatbot based on content you have supplied—Retrieval Augmented Generation—is just one of the ways the API can be used. For the scenario of a domain-knowledge-specific chatbot assistant, conceptually, you start by creating a basic assistant 'shell', where the initialisation parameters supplied select which of OpenAI's baseline Large Language Models (LLMs) the assistant is based on, along with instruction text that is used to scope out the general area of expertise the assistant is to have, along with the style/tone of replies given. Next you specify the Retrieval Tool to be active: this is the built-in sub-system to use if you want to provide your own set of files to augment the knowledge-base the chatbot can draw upon, followed by uploading the files themselves. Dozens of file formats are supported, including textual forms such as Word, PDF, HTML, and commonly used image formats such as PNG and JPEG.

While the set of formats supported is quite extensive, there are also some limits in place regarding the quantity of data that can be given as input to an assistant to provide that domain-specific knowledge. At the time of writing, the number of files that can be supplied was capped at 20, although the individual files themselves

---

[9]https://openai.com/blog/new-models-and-developer-products-announced-at-devday

can be fairly substantive: up to 512MB in size, or else 2 million tokens per file, whichever limit is reached first.

Two core concepts in the Assistants API are Threads and Messages. A Thread is an object type used to represent an ongoing interactive session, such as a particular user's conversation with an assistant. Stored within a given Thread are Messages. This latter object type is used to store both messages sent from the user to an assistant (marked with *role: user*), and vice versa (*role: assistant*). There is also a *system* role message, which can be used to control more central aspects of how a created assistant functions over time.

The capabilities of the Assistants API piqued our interest in the project, raising the question, is it possible to embed a chatbot within the WCDA designed to:

(1) provide a more inquiry-led way for a visitor to engage with the content the digital library provides, suitable for someone who has no prior experience of the venue; and

(2) for a visitor to the archive who has visited The Willow in the past, to respond in a way that helps elicit additional memories?

From a review of the documentation provided by OpenAI about the API, two key areas were identified that required additional software to be written and added to Greenstone. The first was the need to develop a way that the content stored in the digital library could be uploaded as a set of files to be used by the Retrieval Tool. The second was to develop a way in which an interactive text interface using a web-browser-based chatbot could be developed, an aspect that was surprisingly absent in the provided documentation.

To upload a set of files, we co-opted Greenstone's command-line export facility. This allowed us to produce file content in a format suitable for uploading to the Retrieval Tool that integrates document and metadata content together, even if they initially resided in separate files, as source content. There is even a feature to the command-line tool where a set of exported documents can be grouped together into one file, thereby reducing the total number of files generated overall. This provision was actually provided in Greenstone to make more efficient use of the filesystem when the exported data consists of a high number of catalogue records. Regardless of the motivation for the feature, it helped us address the Assistants API's 20-file limit, and so we took the win.

To the existing 'plugouts' for exporting to DSpace, Fedora and the like, we added OpenAIGPTs, selecting HTML as the principal output format for this. With regards to how this works with grouping multiple documents in the digital library into our generated output file (controlled by the *-group_size* parameter when the *export.pl* command is run), we note that HTML, as a format, already encodes structured documents—even for a single document—through elements such as headings *<h1>*, *<h2>*, etc., and more general nesting through elements such as *<div>*. Consequently, we extended this notion in generating an exported HTML file, such that each individual document in the digital library going into a grouped HTML output file was first wrapped up in an *<article>* element.

Next we added the script *create-openai-assistant* to Greenstone. It uses the OpenAI API to create an assistant specific to that collection. Through additional command-line options given to the script, the user creating the assistant can stipulate the name of the assistant
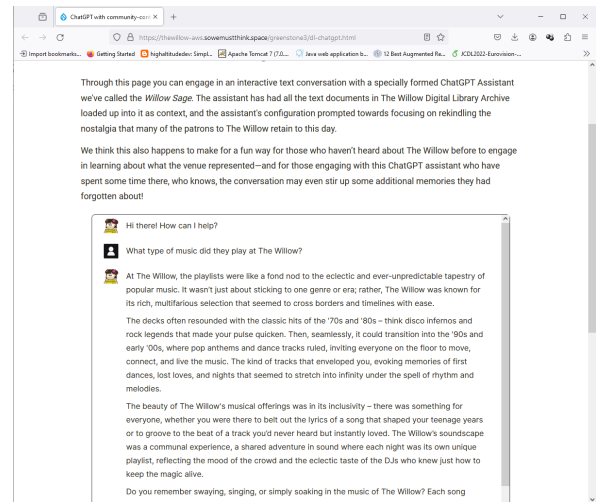

Figure 4: Interacting with 'Willow Sage'.

along with other initialisation settings. In the case of *community-contributions*, we set the assistant's name to 'Willow Sage' and provided an initialisation prompt for the assistant that was geared towards being an 'expert assistant' about this notable venue. The initialisation prompt further stated that the assistant should also seeks to elicit recollections from users interacting with the assistant. The Retrieval Tool subsystem is also activated when the assistant is created, and the file content of the named collection's export directory is read in and uploaded to the assistant.

The second area of code needed related to the development of a web interface through which users could engage with the created 'Willow Sage' assistant. In lieu of OpenAI documentation detailing how to set up a web chatbot connected to one of its models, we surveyed a range of open-source projects, identifying the one called *gpt-4-chat-ui*,[10] which used React for its user-interface components but with minimal other dependencies, as a basis upon which to build. Next we updated the code to work with the version 4 series of the OpenAI software packages so that it could access the Assistants beta API in addition to ChatCompletion, the latter of which only works with the general large-language models provided by OpenAI.

The Assistants API is a significant departure from ChatComplete, providing a fine-grained mechanism to interact with a created Assistant. In addition to the concepts of Threads and Messages, described earlier, there is also a Run class. This allows the code providing the user interface to monitor how the response preparation by the assistant is going, once the latest user message has been submitted. Following the excerpts of code provided in the Assistants API documentation, a workflow was developed that functioned as a replacement for ChatComplete, meaning the React-based web interface could now be directed to connect to an assistant formed by *create-openai-assistant* through the unique identifier assigned to it.

The final step in this second area of development was to style the appearance of the interface to match the design of the main digital-library installation, and integrate it into a page the digital library serves up through an *<iframe/>*. Figure 4 is a snapshot of

---

[10]https://github.com/hillis/gpt-4-chat-ui MIT licensed code

---

Prompt 1: Tell me about The Willow

Ah, The Willow—a venue that was so much more than just a nightclub to those who crossed its threshold. Situated in the heart of York, England, it encapsulated the soul of a community …

Owned by the iconic Tommy and Soo Fong, The Willow wasn't just about the drinks or the dance floor; it was about the experience, the warmth, the sense that no matter who you were outside its doors, inside, you were family. …

The closing of The Willow marked the end of an era. It wasn't just a venue that shut its doors; it felt like saying goodbye to a dear friend for those who'd come to know it as a second home. …

Do you have a memory or a story about The Willow that stands out? It's in sharing these pieces of ourselves that the spirit of such places continues to thrive, long after the last note has faded into the night.

Prompt 2: How did the relationship between the owner Tommy Fong and the DJs affect the music that was played?

Tommy Fong played a significant role in shaping the music policy at The Willow, balancing the DJs' creativity with the venue's musical identity. He had a clear vision for The Willow's atmosphere, desiring to retain its unique vibe by combining modern songs with old classics from The Willow's repertoire. … [1].

This musical direction from Tommy encouraged DJs to experiment and be inventive with their playlists, mixing genres and eras to create a distinctive atmosphere.

Sources

[1] Yes. What it was was that Tommy wanted to keep The Willow … – (from WCA_OH_M04)

Table 2: Example questions posed and prompts entered to 'Willow Sage', along with the (abridged) responses given.

'Willow Sage' in use, responding to the prompt: *What sort of music did they play at The Willow?*

The OpenAI API for Assistants has an Annotate feature, which—when relevant—provides details mapping a section of a returned answer back to the part of a specific source file provided to the Retrieval Tool. Unfortunately, based on postings on the Developers Forum, this part of the API was not yet fully operational (hence the beta-status). However, we liked the intent, and so to compensate for deficiencies, back in the plugout we augmented it to include span tags that explicitly stated which Greenstone document ID the text relates to. This meant that through the *quote* field in a returned Annotation, we were able to identify which document in the digital library collection that part of the response relates to, and from that embed a hyperlink so the user could go from viewing the response given by 'Willow Sage', to seeing the source document in the Digital Library that part of the response was based on.

To give a flavour of the types of responses 'Willow Sage' gives, Table 2 shows the result of consecutive prompts entered, and the responses given (which we have had to edit down in the table to conserve space). While admittedly anecdotal, we were generally impressed with the tone and style of replies generated overall, particularly with respect to rekindling some perhaps forgotten memories someone might have. The phrasing used, however, did have a tendency to skew towards giving a rosy account of The Willow's legacy, with responses given frequently emphasising the inclusive nature of the venue. Entering questions that probed this point of view, such as, 'Did everyone who visited The Willow enjoy their experience?' did give more critical accounts, but even then a follow-up question or two was needed to get to some specifics.

5 DISCUSSION AND CONCLUSIONS

Greenstone's adaptability to new contexts/purposes, as we have demonstrated, makes it especially effective in (re)capturing and pre-serving intangible elements that constitute emotional geographies, especially where infrastructure and resources are lacking (applicable to Global South contexts). For researchers it allows richer possibilities than simply the management and interrogation of qualitative data from interviews and surveys: it facilitates a holistic and iterative process whereby records of all types—archival, audio-visual, interview/survey responses, contributors' reflections—can be brought into open dialogue with each other and juxtaposed in potentially limitless ways. It also encourages users/contributors to become active participants in dynamic narrative building rather than solely passive receivers of 'expert' interpretation, inspiring them to contribute further memories and affective responses that might otherwise have passed undocumented.

Concerning our experiments with ChatGPT, Prompt Engineering (the choice of input texts to AI-based chat systems that specialise it for specific purposes) is an emerging field [16], and this was our first foray into this area. While we were happy overall with the tone used by 'Willow Sage' in its replies, as a consequence of our initialisation instructions (prompts) for the assistant, over time we did go back and make some changes to these. For instance, while we did not specifically instruct 'Willow Sage' to end a reply with a question or two seeking to stimulate additional memories, we could see from our scoping of the initialisation prompts why that had come about. This would be a useful conversational technique without doubt; however, 'Willow Sage' would *always* end its reply in this way, and so we updated its initialisation prompts to give more variety in its endings. In a different situation, we noticed that 'Willow Sage' was using American spelling in its replies. This felt odd, given the subject matter! An additional short sentence directing the assistant to use British spelling was all that was needed to effect this change—something, we note, that contrasts sharply with how such a change would have had to have been programmed using more traditional software-development practices.

All the Greenstone software resources developed for the WCDA, including the site, interface, and the extension from creating assistants based on DL-collection content is available through:

- *https://trac.greenstone.org/browser/gs3-installations/thewillow*
- *https://trac.greenstone.org/browser/gs3-extensions/dl-chatgpt*

ACKNOWLEDGMENTS

REFERENCES

[1] 1973. Oriental Delicacies at New Chinese Restaurant. *Yorkshire Evening Press* (14 June 1973). [Unsigned advertising feature].

[2] 1974. Advertisements in *Nouse*, 12 June 1974 and 10 October 1974; also in *Yorick*, 3, Autumn 1974 (Borthwick Institute for Archives). [*Nouse* and *Yorick* are both student-run publications, University of York].

[3] 2022–24. Recorded interviews conducted by StreetLife researchers with DJs Ken Heald and DJ Max. WCDA.

[4] 2023. Music Venue Trust Annual Report 2023. https://www.musicvenuetrust.com/wp-content/uploads/2024/01/MVT_2023-Annual-Report_Digital.pdf.pagespeed.ce.8gLcgJ--QC.pdf [These statistics only cover Grassroots Music Venues (GMVs), as defined in the report's appendices—the larger category of 'all music venues' could include arenas, concert halls, and so on].

[5] Charlotte Armstrong, Rachel Cowgill, Alan Dix, Christina Bashford, J. Stephen Downie, Mike Twidale, Maureen Reagan, and Rupert Ridgewell. 2021. Towards a Foundation for Collaborative Digital Archiving with Local Concert-Giving Organisations. In *8th International Conference on Digital Libraries for Musicology*. 41–49.

[6] Charlotte Armstrong, Rachel Cowgill, Alan Dix, Christina Bashford, Rupert Ridgewell, Maureen Reagan, Michael Twidale, and J. Stephen Downie. 2023. Reframing Ephemera: Digitisation, Community Music-Making, and Archival Value(s). In *Digital Approaches to Inclusion and Participation in Cultural Heritage*, Danilo Giglitto, Luigina Ciolfi, Eleanor Lockley, and Eirini Kaldeli (Eds.). Routledge, Abingdon, 160–180.

[7] David Bainbridge, Terhi Nurmikko-Fuller, David M. Weigl, and Elizabeth A. Bainbridge. 2023. A Eurovision Digital Library: Let's Open Up with Linked Data. In *2023 ACM/IEEE Joint Conference on Digital Libraries (JCDL)*. IEEE, 301–302.

[8] David Bainbridge and Ian H. Witten. 2020. A renewed look at Greenstone: Entering the Third Decade. *IPSI Transactions on Internet Research* 16, 1 (2020), 3–11.

[9] Sarah Baker (Ed.). 2015. *Preserving Popular Music Heritage: Do-it-Yourself, Do-it-Together*. Routledge, Abingdon.

[10] Gregor Benton and Edmund Terence Gomez. 2008. *The Chinese in Britain, 1800–Present: Economy, Transnationalism, Identity*. Palgrave Macmillan, Basingstoke, 109–131. [This section is on the British Chinese catering trade generally].

[11] Liz Bondi, Joyce Davidson, and Mick Smith. 2007. Introduction: Geography's 'Emotional Turn'. In *Emotional Geographies*, J. Davidson, L. Bondi, and M. Smith (Eds.). Ashgate, Aldershot, 1–16.

[12] Amanda Brandellero, Susanne Janssen, Sara Cohen, and Les Roberts (Eds.). 2014. Special Issue on Popular Music as Cultural Heritage. *International Journal of Heritage Studies* 20, 3 (2014), 219–355. https://doi.org/10.1080/13527258.2013.821624

[13] Sara Cohen, Robert Knifton, Marion Leonard, and Les Roberts (Eds.). 2015. *Sites of Popular Music Heritage: Memories, Histories, Places*. Routledge, Abingdon.

[14] Culture, Media and Sport Committee. 2024. *Seventh Report of Session 2023–24: Grassroots Music Venues*. Report and formal minutes. House of Commons. https://committees.parliament.uk/publications/44704/documents/222242/default/

[15] Alan Dix, Elizabeth Jones, Rachel Cowgill, Charlotte Armstrong, Rupert Ridgewell, Michael Twidale, J. Stephen Downie, Maureen Reagan, Christina Bashford, David Bainbridge, Carys-Ann Neads, and Vince Davies. 2023. Enriching Cultural Heritage Communities: New Tools and Technologies. *Interacting with Computers* (2023). [In press].

[16] Sabit Ekin. 2023. Prompt Engineering for ChatGPT: A Quick Guide to Techniques, Tips, and Best Practices. *TechRxiv* (May 2023). https://doi.org/10.36227/techrxiv.22683919.v2

[17] Victoria Finan. 2015. Farewell, the Willow - the Berghain of York. *The Guardian* (28 July 2015). https://www.theguardian.com/uk-news/the-northerner/2015/jul/28/farewell-the-willow-the-berghain-of-york

[18] Vicki Fong. 2023–24. Interviews with Fong family members and personal recollections. WCDA curated video stories. [ThomFong].

[19] Phil Hubbard. 2007. The Geographies of 'Going Out': Emotion and Embodiment in the Evening Economy. In *Emotional Geographies*, J. Davidson, L. Bondi, and M. Smith (Eds.). Ashgate, Aldershot, 117–134.

[20] Michael Kay. 2008. *XSLT 2.0 and XPath 2.0 Programmer's Reference* (4th ed.). Wiley, Indianapolis, IN.

[21] Mathew A. Logan and Chern Li Liew. 2023. GLAM Convergence Revisited: An Examination of User Perception and Experience. *Journal of Library Administration* 63, 8 (2023), 1014–1043.

[22] Liam Maloney and John Schofield (Eds.). 2021. *Music and Heritage: New Perspectives on Place-making and Sonic Identity*. Routledge, Abingdon.

[23] Ian H. Witten, David Bainbridge, and David M. Nichols. 2009. *How to Build a Digital Library* (2nd ed.). Morgan Kaufmann, San Francisco, CA, USA.

[24] Veronika Zink. 2019. Affective Communities. In *Affective Societies: Key Concepts*, Jan Slaby and Christian von Scheve (Eds.). Routledge, Abingdon, 289–299.