Digital humanities and digitised cultural heritage

Citation for published version:
Terras, M 2022, Digital humanities and digitised cultural heritage. in J O'Sullivan (ed.), *The Bloomsbury Handbook to the Digital Humanities.* 1 edn, Bloomsbury Handbooks, Bloomsbury Academic, pp. 255-266. https://doi.org/10.5040/9781350232143.ch-24

Digital Object Identifier (DOI):
10.5040/9781350232143.ch-24

Published In:
The Bloomsbury Handbook to the Digital Humanities

CHAPTER TWENTY-FOUR

Digital Humanities and Digitized Cultural Heritage

MELISSA TERRAS (UNIVERSITY OF EDINBURGH)

The Gallery, Library, Archive, and Museum (GLAM) sector have been involved in a concerted—if not evenly distributed—effort to mass-digitize collections for over thirty years. Over this period, digitization processes and best practices have been established, and there have been various phases and mechanisms of delivery which now see a vast, if patchy, digital cultural heritage landscape. Throughout, the Digital Humanities (DH) community has engaged with digitized primary historical sources, using the *products* of the digitization *process* for research, teaching, and engagement. The DH community has built corpora, infrastructure, tools, and experimental interventions upon digitized cultural material, while also computationally mining and examining content. This chapter will analyze and synthesize the different activities those in DH undertake with digitized cultural heritage, stressing that those wishing to utilize it in research and teaching must engage with digitization processes as well as the product to best understand the data's epistemic foundations. Developing a holistic understanding of the digitization environment that provides the data DH activities are built upon can allow DH to contribute to it in return, influencing and impacting future digitization activities. This chapter sketches out an agenda for the Digital Humanities in considering digitized content, while also reflecting on how best to create a feedback loop to undertake activities and produce research that will be of interest to digitization providers. By working together with collections and their owners, DH activities can help construct a roadmap towards building inclusive digital heritage datasets that are useful, re-usable, and point to the benefits of user engagement with our digitized past, while also more fully understanding and influencing the processes which create digital resources, and the wider digital scholarship landscape.

DIGITIZATION IN THE GLAM SECTOR

Since the 1970s, GLAM have been undertaking digitization, using the affordances of digital networked infrastructure to improve the management of, increase engagement with, and provide access to heritage collections (Hughes 2004; Parry 2013; Terras 2011). Mass-digitization began in earnest in the 1990s (Lee 2001, 160). An increase in digital activity across society as a whole has meant that users now expect heritage to also be available in digital form (Falk and Dierking 2016, 122), with a related increase in digital interactive and immersive technologies in heritage contexts (Pittock 2018, 5). The technical aspects of routine, service, or mass-digitization are now mostly

a solved problem, with best practice guidelines available for others to consult.[1] There are expert communities of practice across the heritage sector regarding data acquisition and recording.[2] The importance of digitization to the heritage landscape was highlighted by the Covid-19 epidemic, when many institutions worldwide had to restrict physical access: in response, 86 percent of museums increased their online presence and/or the amount of content they were placing online (ArtFund 2020), with emerging opportunities regarding the reframing of digitized content as an essential part of cultural memory (Kahn 2020).

What is the relationship of the DH to this digitization landscape? How best can researchers respond to it, not only to use the *products* of heritage digitization, but to engage in the *process*, to encourage inclusion and engagement? Most obviously, DH techniques such as text and network analysis are dependent on machine-processable data, usually generated from the digitization of primary historical sources.[3] These data sources may be small scale, and possibly creatable by a lone researcher undertaking digitization themselves for a particular project. Both the copyright exceptions that allow digital copies of material to be created for nonprofit data-mining research purposes, and the advancing technological infrastructures mean that, even with a modern smartphone and a few choice apps, high-quality data regarding historical collections can now be efficiently created for later processing and analysis (Rutner and Schonfeld 2012). However, more often than not, researchers use the products of digitization carried out by a GLAM institution, or a digitization service provider working with material from a GLAM collection.

The choice of which materials are selected for digitization may be driven by the needs of a particular research project, with allied external funding. Nowadays, in a changing resource landscape, it is becoming more common for GLAM institutions to be expected to provide their own internal budgets for digitization. Researchers may build questions around data from primary sources that were pre-emptively digitized as they were believed by GLAM curators and their institutions to be critical material, in the hope that if you build it they will come (Warwick et al. 2008). Institutions may also engage with their research communities to understand how best to triage and schedule digitization of their collections (Wingo, Heppler, and Schadewald 2020). Alternatively, DH researchers can sometimes access content that was mass-digitized because there was a suitable business case for monetizing it, for example, local newspapers that were primarily digitized by publishers for paid-for access by the online genealogy community (Hauswedell et al. 2020).

The products of digitization have changed over the past thirty years, and with that, so has the way DH activities interact with digitized content. There initially was naivety and techno-futurist excitement surrounding digitization (which saw a lot of material subject to "scan-and-dump," where large collections were placed online although only accessible one item at a time, and there was a hope that people would find this transformative (Warwick et al. 2008)). Those in the Digital Humanities have always been interested in re-using available content, and ascertaining how these platforms were being used by others, to best frame how future resources may be deployed. The mid- to late 2000s saw the rise of Museum 2.0, with a growth in participatory culture and co-creation, aligned with providing content via a growing range of social media platforms (including Facebook, Twitter, Tumblr, Instagram, and more recently SnapChat, and TikTok) as well as data and information sharing platforms (such as Flickr Commons and Wikipedia). Research demonstrated that this type of activity both reached and broadened GLAM audiences (Finnis 2011; Hughes 2012). Over the past decade, digital cultural heritage aggregators, such as Europeana,[4] Google Arts and Culture,[5] the Internet Archive,[6] and the Digital Public Library of America[7] have

become the central point of call for a user wishing to browse cultural or heritage content, and attention has been paid to the politics and biases that are embedded within these infrastructures and mass-digitization programs (Thylstrup 2019; Kizhner et al. 2020). There has also been interest in how to extract further information from digitized heritage: mechanisms for users to engage with digitized resources in order to complete more complex tasks such as identification, cataloging, and transcription via crowdsourced volunteering, are now in the mature phase of delivery and study by those in DH (Ridge 2014; Terras 2016). An increasing range of approaches to produce digital scholarly editions that build upon digitized images of texts are now used (Bleier et al. 2018). Over the past decade, the OpenGlam[8] movement has made these types of activities easier, providing frameworks, pilot studies, and use-cases for institutions aiming to provide "ethical open access to cultural heritage" (OpenGlam n.d.) by releasing resources onto the open web with a range of licenses (usually via the Creative Commons[9] framework). Moving beyond the relationship of a user encountering one heritage item in digital form at a time, to encountering and synthesizing many, the "Collections as Data" movement, where large-scale data deposits of complete collections are made available under open licensing to encourage users to take, manipulate, analyze, and reframe digital collections (Padilla 2018), is of much interest to many in the DH. As technologies that can produce machine-processable data have reached maturity, such as handwritten text recognition (Muehlberger et al. 2019), and voice to text speech recognition (Picheny et al. 2019), the volume of data derived from GLAM material available for scholars is rapidly expanding, and requires understanding in best practice for both GLAM institutions in creating, hosting, delivering, and stewarding content, and DH approaches in utilizing it at this scale.

However, it should always be remembered that although digitization is the starting point for these DH activities only a fraction of content has been digitized. While 82 percent of GLAM institutions across Europe have a digital collection or are engaged in digitization (Nauta et al. 2017, 5) after thirty years of large-scale investment in digital, on average 22 percent of heritage collections across Europe have been digitized, and only 58 percent of collections have even been cataloged in a collections database (Nauta et al. 2017, 6). There is still much to do, then, in both digitizing, understanding how best to deliver and use digitized content in research and teaching, and extracting further value from digitized content (Terras et al. 2021). The DH community is well placed to both utilize the *products* of such activities, and advise and become integrated into the *processes*, questioning the selection choices that are made along the way, and highlighting the impacts that these have on diversity and inclusion, while also being able to use digital tools to reflect on the make-up, history, and constitution of collections which steward our cultural heritage.

ENGAGING WITH THE DIGITIZATION PROCESS

GLAM digitization processes and resources do not exist solely to provide data for further analysis by those within DH. Established reasons given for mass-digitization of heritage content include: increasing access to wider communities; supporting preservation; collections development; raising the profile of collections and institutions; while also supporting research, education, and engagement (Hughes 2004, 8–17). There is nowadays much focus on engagement and its related metrics in digital GLAM strategies, which is partly driven by funders and the quantified nature of feedback regarding online estates. For example, in the UK, the Department for Digital, Culture, Media and Sport launched a "Culture is Digital" policy paper in 2018 that frames the best use of technology

in GLAM as driving "audience engagement, boosting the digital capability of cultural organizations and unleashing the creative potential of technology." The effect is that many organizations in the sector are, by necessity, focusing rather shallowly on reporting on clicks and likes, "using digital technology to engage audiences" (DCMS 2018). This is at the expense of thinking more holistically about digital strategies that can benefit both the organization and their communities in various ways (Finnis 2020).

Digitization is just part of a suite of digital activities transforming the GLAM sector (including "Strategy and Governance; Cultural Programme; Places and Spaces; Marketing & Communications; HR; IT; and Finance & Operations," see Digital Compass (2020) for a toolkit designed to assess and expand an organization's use of digital). Amongst this changing landscape, it has been noted that it is not clear how GLAM organizations ever intended their audiences to re-use mass-digitized heritage content (Harrison et al. 2017). Additionally, "institutions do not have the resources of major technology providers for research and development, and are not as easily able to recover from costly technological failures in design, delivery, or public relations" (Terras et al. 2021). As a result, making the most of digitized cultural heritage data for DH research purposes (or an approach now more commonly known as "Digital Scholarship" within the library and heritage sector (Greenhall 2019)) requires sensitivity from scholars who often have to navigate very different infrastructures and processes, institutional focus, and available resources. Equipping a GLAM institution "for digital scholarship, navigating a world of Big Data, and understanding the implications of this shift, requires organisation-wide re-evaluation" (Ames and Lewis 2020), at a time of much austerity for the sector. There is therefore an opportunity for DH to contribute to this space, working alongside the creators of digitized content, in establishing how best to allow researchers to gain access to content, and undertake the type of advanced digital methods that are espoused by DH.

DIGITIZATION AND THE DIGITAL HUMANITIES AGENDA

DH researchers have much to contribute in understanding, quantifying, critiquing, elucidating, and contributing to both the process and products of digitization, closing the gap between institutions and researchers in this data-led space. GLAM institutions themselves are grappling with the challenges that the volume of digitized content they create now presents, given that "new services, processes and tools also need to be established to enable these emerging forms of research, and new modes of working need to be established to take into account an increasing need for transparency around the creation and presentation of digital collections" (Ames and Lewis 2020). How best should the DH respond to this? In their analysis of "Big Data Stakeholders," Zwitter (2014) identifies three categories: collectors, utilizers, and generators, noting that there are "power relationships" between the three, and that the role of the collector is essential in determining what is collected and made available to others. Therefore, an intervention is needed from DH, to engage as and with collectors, utilizers, and generators, bridging and reframing the relationship of GLAM as provider and scholar as consumer of digitized content. Digital humanists should be responding to this call to arms by not only passively using GLAM content, but in producing research and other digital outputs that engage in discussions with the GLAM sector regarding digitization processes, deployment, engagement, and adoption.

What, then, should be the DH agenda, in the digitization space? Points of meaningful tension and possible contribution exist in: keeping abreast of technological developments and possible deployment in the GLAM sector; understanding institutional, political, and financial drivers of digitization, and the legal frameworks which bind and contain activities, to understand how these affect availability and usefulness of its products; making best use of the mass of digitized content at scale now emerging from the sector; and understanding how social media, and other external digital platforms, can best be used to promote engagement and inclusion.

Although the technical processes of routine digitization are now understood,[10] there are always new developments which require reception studies to understand how they may be best used within the GLAM sector (as well as interdisciplinary projects to be undertaken with the computational and engineering sciences in developing advanced capture and processing techniques). Currently, there is much interest in applying machine-learning approaches to digitized content (Jaillant 2021), capture techniques such as hyper-spectral imaging (Howell and Snijders 2020), and the implications regarding the recent affordability of 3D scanning techniques such as LiDAR (Collins 2018). Additionally, the Foundation Projects of the UK's "Towards a National Collection"[11] are investigating current best practices, presenting roadmaps for required research in: cross-walking catalogs and other digitized content sources;[12] best practice in sharing digitized images via IIIF,[13] implementing a system of Persistent Identifiers that can provide long-lasting references to digital heritage objects;[14] utilizing computer vision and deep-learning methods to identify patterns and linkages across collections using image processing;[15] making best use of geospatial data to connect diverse collections;[16] preserving and sharing born-digital and hybrid objects;[17] understanding the best way to engage with volunteers in citizen research projects;[18] and developing ethical methodologies and technical approaches to address outmoded and offensive racist, sexist, colonial, and imperial language in digital heritage information landscapes.[19] Those in the DH community are well placed to work with a broad range of collaborators in these many areas to establish how developing technologies will impact, and could be efficiently deployed, in a GLAM context, influencing DH research further down the line regarding how we teach and research with these techniques and their data.

However, as well as the technical aspects to creation of digitized content, DH could do much to elucidate the context regarding *which* digital resources are created within the GLAM sector. Choices made about digitization are a matter related to social justice: if we can only analyze historical and archival information that has been digitized, it is worth paying closer attention to the politics of selection, and how that relates to issues of diversity and inclusion (Thylstrup 2019; Zaagsma 2019). There are various tensions at play when decisions are made about resource allocation: cultural heritage organizations are inherently risk-averse, with often conflicting requirements from funders, a board of trustees, senior management team, curatorial staff, conservators, and digitization professionals. The sector, which is moving into post Covid-19 operationalizing, is still reeling after years of austerity, and resources for digitization and related digital activities are scarce. There are ethical dimensions regarding digitization funding: "the availability of funding, public or private, plays a key role in enabling digitization projects in the first place … Funding is … influenced by … memory politics and the way in which a given country's or group's past, or aspects thereof, resonate in public discourses and debates" (Zaagsma 2019). For example, what, pray tell, could possibly be the current political drivers for government funding for a new digital "National Collection: Opening UK Heritage to the World" in recently post-Brexit Britain? (AHRC 2021).

How will the choices made in this investment impact what we can access, research, and study, for years to come? Furthermore, as standard digitization processes continue to scale, and given cultural heritage institutions have limited resources, the relationship with technology providers and major publishers, and the constraints and opportunities that arise with regard to commercial digitization of heritage, is worthy of further thought and consideration (Thylstrup 2019; Hauswedell et al. 2020). All of these aspects are worthy of appraisal and investigation by DH scholars when using the products of digitization processes, which will be of use to a broad constituency, from museum studies, to library science, to critical heritage studies, to DH itself. Only in understanding the choices made in the creation of digitized content, can it be fully appraised and utilized.

Attention is also needed regarding the legal frameworks which bound and contain digitization activities, and therefore shape the online GLAM environment. Copyright is a major barrier to digitization, particularly around restrictions placed on the digitization of modern in-copyright materials (Kariyawasam and Adesanya 2019) and "orphan works," the vast majority of cultural heritage holdings where the copyright owner of an item cannot be identified, which has ramifications for choices made around digitization (Korn 2009). Other legal frameworks such as Intellectual Property Rights, and General Data Protection Regulations, likewise affect digitization selection. Understanding how these legal frameworks (and the business models that grow from them, see Terras et al. (2021)) intersect with risk-averse institutions and the digitization process will help understand—and improve—the digitized cultural heritage environment.

Understanding GLAM collections and material at scale is an emerging research area. There is much potential for DH in the appraisal of scale, focus, and topic of previously digitized content, as well as elucidating biases contained within them. In this current phase of digitization, we are realizing that the choices previously made over the past four decades have ethical and often colonial aspects, and that there is a duty of care that comes with making cultural and heritage collections available to a wider audience (Odumosu 2020). These biases play out at scale, with unforeseen consequences for audiences and researchers (Kizhner et al. 2020). There is much more that can be done with digitized cultural heritage to foreground and support minority voices within collections, in order to "mainstream equality," for example around aspects of gender, race, or LGBTQ+ issues (National Galleries Scotland 2019). The DH community is well placed to analyze and synthesize "Collections as Data" (Padilla 2018), understanding the substance, and biases of cultural heritage collections, as well as searching and analyzing them. Catalog metadata itself is a crucial but overlooked information resource that is ripe for mining, visualization, and further analysis to understand collection structure, coverage, themes, and lacunae (Lavoie 2018). Additionally, the recursive process of applied data science means that new questions that *could* be asked regarding institutional collections are sure to emerge, and understanding and making these explicit would be a contribution from DH research, itself.

Digitized cultural heritage often intersects with social media (in GLAM policies as well as activities). This is a fascinating space, as the first point of contact with heritage with a wide audience is now often over these distributed channels. The specific institutional context, resources, opportunities, struggles, and approach to risk management, is all set out for public display (Koszary 2018), and resulting analysis. For example, the Digital Footprints and Search Pathways[20] project is researching the effect of the Covid-19 pandemic on the audiences of online collections, to help GLAM institutions develop engagement plans and policies for crisis situations. There is much ongoing work to be done in DH classes and projects around the intersection of digitized heritage and ever-changing online platforms.

IMPLEMENTING THE DIGITAL HUMANITIES DIGITIZATION AGENDA

How practically should the DH community undertake activities that center digitization, to engage with, feedback to, and improve it? Firstly, it is necessary to ask of any dataset derived from historic primary resources how it came into being—how, where, and why was it digitized? The flip-side of this is to ask, how, where and what was *not* digitized that may be aligned to the research question, or would increase accessibility and inclusion if it were to be made available in machine-processable form. This should be part of the discussion and presentation of any DH project that depends and builds upon digitized material, so that the reader can fully understand the limitations and possibilities framing the undertaking.

Secondly, DH teaching programs should also be integrating the study of digitization processes into the curriculum, gaining both practical technical experience, and understanding application and theory.[21] This is another skill set that will allow our graduates to go on to find useful and gainful employment, where their critical and technical skills can contribute to the "digitalisation of society" (Bowen and Giannini 2014), but from a different intellectual tradition than the computational sciences (which in itself is a matter of diversity and inclusion within technology industries (Lachman 2018)). The processes and products of digitization of cultural heritage are mirrored in current efforts to digitize medical, banking, educational, legal, and judicial records (which can have social consequences: see for example Favish (2019) on the effect of using OCR to release the backlog of Deeds Registry records in post-apartheid South Africa). There is much a DH lens can introduce into learning about digitization processes and how they overlap with wider societal concerns.

Thirdly, given the scale of DH relevant digitized content is rapidly expanding (particularly with transcripts automatically generated from manuscript material via HTR, or from audio material via speech recognition, and from the Collections as Data movement), skill sets are needed to be able to best use this data. Developing best practice approaches to historical text and data mining at scale should be an aspiration for DH (including learning and applying prior work from related fields such as statistics, Natural Language Processing, and High Performance Computing, or working in conjunction with individuals from those domains who have the skill sets to methodically tackle data at scale). As ever, it is difficult to recommend one particular tool, methodology, or method that represents "Digital Humanities"—the selection will be entirely dependent on the research question being asked and the dataset it is being applied to. However, there is much innovation happening in this space, for example Tim Sherratt's *GLAM Workbench*[22] tools, which provide a set of Jupyter Notebooks for exploration and re-use of data from cultural heritage institutions. For example, in working with data from Trove[23] (which aggregates collections from Australian GLAM providers) there are notebooks for visualizing: searches in Trove's digitized newspapers; the total number of newspaper articles in Trove by year and state; the geographical spread of content mapped by place of publication in Trove; and to analyze rates of OCR (Optical Character Recognition), amongst others.[24] Another set of crucial resources for learning this type of processing is the Programming Historian[25] which hosts rigorously peer-reviewed tutorials for digital research skills, which are openly licensed for re-use. The Data Carpentries[26] lesson programs, which teach foundational coding and data science skills, are also a resource to be drawn down on. In learning what is possible via these approaches, scholars can respond best to the opportunities presented by the growing body of machine-processable GLAM data. DH researchers should also keep a

close eye on the developing Collections as Data community and their datasets; for example, the Data Foundry[27] at the National Library of Scotland has made available large-scale, openly licensed digitized collections with transparent documentation for others to build upon, which now underpin the Library Carpentry Text and Data Mining[28] lesson. There are many cultural heritage institutions moving in this direction.

Fourthly, in developing skills more commonly associated with information rather than computer scientists, such as qualitative interview, survey, and observation skills (Gorman and Clayton 2005), as well as user experience (UX) design skills (Kuniavsky 2010), DH as a field will be able to ask broader questions about the creation, usability, use, and impact of digitized resources, particularly within their chosen research area (such as analyzing social media activity). In doing these, DH scholars will be able to gauge the "epistemic affordances" of digitized content: the abilities, possibilities, and limitations of this environment when used in knowledge creation (Markauskaite and Goodyear 2017). Qualitative methods will be useful here, including action research (Schön 1983), grounded theory (Strauss and Corbin 1997), and content analysis (Krippendorff 2018) to synthesize different reports or analyze public-facing information. These are useful options to add to and expand the DH curriculum, as well as being crucial research skills where navigating complex information about decision-making environments with intersecting power structures, in systematic and rigorous ways.

Fifthly, it is also probable that these different skill sets need different individuals working together as part of a project team to undertake this type of research (interdisciplinary collaborative work being a perceived norm in many DH projects (Klein 2015)). It is worth stressing that information and museum professionals including librarians, curators, conservators, social media managers, and digitization practitioners have much expertise that can both benefit DH projects and allow professional development and opportunities for DH scholars: this is never one-way skills transfer. Individuals within DH have a lot to learn from our professional collaborators.

Finally, to ensure the impact of activities in this space, this crucial research should be published in venues that interact with a wider audience rather than the core DH publication and conference venues. (I have always wondered why, if DH is so interdisciplinary, so many of its practitioners mainly present their work in DH silos, rather than publishing in overlapping venues, where librarians, curators, digitization experts, heritage professionals, information and computational scientists, and wider industry contacts will more easily come across it.) Such venues are not hard to find, and many have overlaps with the Digital Humanities conference,[29] such as iConference,[30] or Museums and the Web.[31] Likewise, museum, critical heritage, library and information science, and digital sociology journals are all apposite venues for DH researchers to publish in, that will allow others grappling with building the digital scholarship landscape to become more aware of the types and range of extant and future DH activities.

CONCLUSION

This chapter has sketched a forward-facing agenda for the Digital Humanities community regarding how it perceives, considers, uses, analyzes, and reports on activities with digitized cultural heritage content. In doing so, it has made concrete recommendations regarding how digitized content can continue to underpin while also be considered part of core Digital Humanities research and teaching activities. There is much useful activity still to be done at the nexus of DH and GLAM digitization

practices. In doing so, the DH community will be able to inform and work with GLAM institutions in best responding to, delivering, and guiding the future of digitized content. In navigating together the "new challenges around data management, storage, rights, formats, skills and access" that these mass-digitized resources present (Ames and Lewis 2020), the DH community can contribute to the wider, emerging digital scholarship landscape.

NOTES

1. http://www.digitizationguidelines.gov.
2. See for example the Association for Historical and Fine Art Photography (https://ahfap.org.uk) which has an active mailing list, or the ImageMuse discussion group for cultural heritage imaging (http://www.imagemuse.org).
3. An alternative to this is utilizing born-digital content (Kirschenbaum 2013): we concentrate on digitized heritage content within this chapter to allow focus.
4. Europeana, https://www.europeana.eu/en.
5. Google Arts and Culture, https://artsandculture.google.com.
6. Internet Archive, https://archive.org.
7. Digital Public Library of America, https://dp.la.
8. OpenGlam Network, https://openglam.org.
9. Creative Commons, https://creativecommons.org.
10. Federal Agencies Digital Guidelines Initiative, http://www.digitizationguidelines.gov.
11. Towards A National Collection, https://www.nationalcollection.org.uk/projects.
12. Heritage Connector, Science Museum, UK, https://www.sciencemuseumgroup.org.uk/project/heritage-connector/.
13. Practical applications of IIIF as a building block towards a digital National Collection, National Gallery, UK, https://tanc-ahrc.github.io/IIIF-TNC/ and https://iiif.io.
14. Towards a National Collection – HeritagePIDs, British Library, UK, https://tanc-ahrc.github.io/HeritagePIDs/.
15. Towards a National Collection – Deep Discoveries, National Archives, UK, https://tanc-ahrc.github.io/DeepDiscoveries/.
16. Locating a National Collection, British Library, UK, https://tanc-ahrc.github.io/LocatingTANC/.
17. Preserving and sharing born-digital and hybrid objects from and across the National Collection, Victoria & Albert Museum, UK, https://gtr.ukri.org/projects?ref=AH%2FT01122X%2F1.
18. Engaging Crowds: Citizen research and heritage data at scale, National Archives, UK, https://tanc-ahrc.github.io/EngagingCrowds/.
19. Provisional Semantics, Addressing the challenges of representing multiple perspectives within an evolving digitised national collection, Tate, UK, https://www.tate.org.uk/about-us/projects/provisional-semantics.
20. Digital Footprints and Search Pathways, Centre for Research in Digital Education, University of Edinburgh, https://www.de.ed.ac.uk/project/digital-footprints-and-search-pathways.
21. Where this isn't done already: some courses, such as the MA/MSc in Digital Humanities in the Information Studies department at UCL (established by the author in 2015, forgive me) have a practice-led "Introduction to Digitisation" module, that teaches the principles and practices of the digitization life cycle in a hands-on environment (UCL 2021).
22. GLAM Workbench, https://glam-workbench.github.io/about/.
23. Trove, National Library of Australia, https://trove.nla.gov.au.
24. GLAM Workbench, Trove Newspapers, https://glam-workbench.github.io/trove-newspapers/.
25. Programming Historian, https://programminghistorian.org.
26. Data Carpentry, https://datacarpentry.org.
27. Data Foundry, National Library of Scotland, https://data.nls.uk.

28. Library Carpentry: Text and Data Mining, http://librarycarpentry.org/lc-tdm/.
29. Digital Humanities Annual Conference, Alliance of Digital Humanities Organisations, https://adho.org/conference.
30. iSchools Annual Conference, iSchools, https://ischools.org/iConference.
31. Museums and the Web Annual Conference, https://www.museweb.net/conferences/.

REFERENCES

Ames, Sarah and Stuart Lewis. 2020. "Disrupting the library: Digital scholarship and Big Data at the National Library of Scotland." *Big Data & Society* 7 (2): 2053951720970576.

ArtFund. 2020. "Covid-19 Impact: Museum Sector Research Findings. Summary Report." https://www.artfund.org/assets/downloads/art-fund-covid19-research-report-final.pdf.

Arts and Humanities Research Council. 2021. "Towards a National Collection: Opening UK Heritage to the World." https://ahrc.ukri.org/research/fundedthemesandprogrammes/tanc-opening-uk-heritage-to-the-world/.

Bleier, Roman, Martina Bürgermeister, Helmut W. Klug, Frederike Neuber, and Gerlinde Schneider, eds. 2018. *Digital Scholarly Editions as Interfaces*, vol. 12. BoD–Books on Demand.

Bowen, Jonathan P. and Tuia Giannini. 2014. "Digitalism: The New Realism?" *Electronic Visualisation and the Arts (EVA 2014)*: 324–31.

Collins, Lori D. 2018. "Terrestrial Lidar." *The Encyclopedia of Archaeological Sciences*. Wiley Online Library. https://doi.org/10.1002/9781119188230.saseas0575.

Department for Digital, Culture, Media and Sport. 2018. "Culture is Digital." https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/687519/TT_v4.pdf.

Digital Culture Compass. 2020. "Tracker: Quick Guide." https://digitalculturecompass.org.uk/using-the-tracker/.

Falk, John H. and Lynn D. Dierking. 2016. *The Museum Experience Revisited*. New York, NY: Routledge.

Favish, Ashleigh. 2019. "Data Capture Automation in the South African Deeds Registry using Optical Character Recognition (OCR)." Master's thesis, Faculty of Commerce, African Institute of Financial Markets and Risk Management. http://hdl.handle.net/11427/31389.

Finnis, Jane. 2011. "Let's Get Real 1: How to Evaluate Success Online." Culture24, Brighton, UK. https://www.keepandshare.com/doc/3148918/culture24-howtoevaluateonlinesuccess-2-pdf-september-19-2011-11-15-am-2-5-meg?da=y&dnad=y.

Finnis, Jane. 2020. "The Digital Transformation Agenda and GLAMs. A Quick Scan Report for Europeana." Culture24, July. https://pro.europeana.eu/files/Europeana_Professional/Publications/Digital%20transformation%20reports/The%20digital%20transformation%20agenda%20and%20GLAMs%20-%20Culture24%20findings%20and%20outcomes.pdf.

Gorman, Gary Eugene and Peter Robert Clayton. 2005. *Qualitative Research for the Information Professional: A Practical Handbook*. London: Facet Publishing.

Greenhall, Matt. 2019. "Digital Scholarship and the Role of the Research Library. The Results of the RLUK Digital Scholarship Survey." Research Libraries UK. https://www.rluk.ac.uk/wp-content/uploads/2019/07/RLUK-Digital-Scholarship-report-July-2019.pdf.

Harrison, Rodney, Hana Morel, Maja Maricevic, and Sefryn Penrose. 2017. "Heritage and Data: Challenges and Opportunities for the Heritage Sector." Report of the Heritage Data Research Workshop, British Library, London, June 23. https://heritage-research.org/app/uploads/2017/11/Heritage-Data-Challenges-Opportunities-Report.pdf.

Hauswedell, Tessa, Julianne Nyhan, Melodee H. Beals, Melissa Terras, and Emily Bell. 2020. "Of Global Reach yet of Situated Contexts: An Examination of the Implicit and Explicit Selection Criteria that Shape Digital Archives of Historical Newspapers." *Archival Science*: 139–65. https://doi.org/10.1007/s10502-020-09332-1.

Howell, David and Ludo Snijders. 2020. "Spectral Imaging." In *Conservation Research in Libraries*, 172–96. Current Topics in Library and Information Practice. De Gruyter Saur

Hughes, Lorna M. 2004. *Digitizing Collections: Strategic Issues for the Information Manager*.London: Facet Publishing.

Hughes, Lorna M., ed. 2012. *Evaluating and Measuring the Value, Use and Impact of Digital Collections*. London: Facet Publishing.

Jaillant, Lisa, ed. 2021. *Archives, Access and Artificial Intelligence: Working with Born-Digital and Digitised Archival Collections*. Berlin: Transcript.

Kahn, Rebecca. 2020. "Locked Down Not Locked Out – Assessing the Digital Response of Museums to COVID-19." *LSE Impact of Social Sciences Blog*, May 8. https://blogs.lse.ac.uk/impactofsocialsciences/2020/05/08/locked-down-not-locked-out-assessing-the-digital-response-of-museums-to-covid-19/.

Kariyawasam, Kanchana and Olumayowa O. Adesanya. 2019. "Impact of Copyright Law on Mass Digitisation." *New Zealand Business Law Quarterly* 25 (1): 46–68.

Kirschenbaum, Matthew. 2013. "The.txtual Condition: Digital Humanities, Born-Digital Archives, and the Future Literary." *Digital Humanities Quarterly* 7 (1). http://www.digitalhumanities.org/dhq/vol/7/1/000151/000151.html.

Kizhner, Inna, Melissa Terras, Maxim Rumyantsev, Valentina Khokhlova, Elisaveta Demeshkova, Ivan Rudov, and Julia Afanasieva. 2020. "Digital Cultural Colonialism: Measuring Bias in Aggregated Digitized Content Held in Google Arts and Culture." *Digital Scholarship in the Humanities* 36 (3): 607–40.

Klein, Julie T. 2015. *Interdisciplining Digital Humanities: Boundary Work in an Emerging Field*. Minneapolis, MN: University of Michigan Press.

Korn, Naomi. 2009. "In from the Cold: An Assessment of the Scope of Orphan Works and its Impact on the Delivery of Services to the Public." JISC Content, April. https://naomikorn.com/wp-content/uploads/2020/09/SCA_CollTrust_Orphan_Works_v1-final.pdf.

Koszary, Adam. 2018. "Seven Broad Statements that May or May Not Help Your Museum Do a Bit Better at Social Media." November 12. https://medium.com/@adamkoszary/seven-broad-statements-that-may-or-may-not-help-your-museum-do-a-bit-better-at-social-media-6173c3b3af0c.

Krippendorff, Klaus. 2018. *Content Analysis: An Introduction to Its Methodology*, 4th edn. London: SAGE Publications.

Kuniavsky, Mike. 2010. *Smart Things: Ubiquitous Computing User Experience Design*. Burlington, MA: Elsevier.

Lachman, Richard. 2018. "STEAM Not STEM: Why Scientists Need Arts Training." The Conversation, January 17. https://theconversation.com/steam-not-stem-why-scientists-need-arts-training-89788.

Lavoie, Brian. 2018. "How Information about Library Collections Represents a Treasure Trove for Research in the Humanities and Social Sciences." *LSE Impact of Social Sciences Blog*, September 19. https://blogs.lse.ac.uk/impactofsocialsciences/2018/09/19/how-information-about-library-collections-represents-a-treasure-trove-for-research-in-the-humanities-and-social-sciences/.

Lee, Stuart D. 2001. *Digital Imaging: A Practical Handbook*. London: Facet Publishing.

Markauskaite, Lina, and Peter Goodyear. 2017. "Epistemic Thinking." In *Epistemic Fluency and Professional Education*, 167–94. Dordrecht, Springer.

Muehlberger, Guenter, Louise Seaward, Melissa Terras, et al. 2019. "Transforming Scholarship in the Archives through Handwritten Text Recognition: Transkribus as a Case Study." *Journal of Documentation* 75: (5). https://www.emerald.com/insight/content/doi/10.1108/JD-07-2018-0114/full/html.

Nauta Gerhard Jan, Wietske van den Heuvel, and Stephanie Teunisse. 2017. D4.4 Report on ENUMERATE Core Survey 4. Europeana DSI 2- Access to Digital Resources of European Heritage, Europeana. https://pro.europeana.eu/files/Europeana_Professional/Projects/Project_list/ENUMERATE/deliverables/DSI-2_Deliverable%20D4.4_Europeana_Report%20on%20ENUMERATE%20Core%20Survey%204.pdf.

National Galleries Scotland. 2019. "Mainstreaming Equality at the National Galleries of Scotland." April. Edinburgh: National Galleries of Scotland. https://www.nationalgalleries.org/sites/default/files/features/pdfs/Mainstreaming_Equality_Report_April_2019.pdf.

Odumosu, Temi. 2020. "The Crying Child: On Colonial Archives, Digitization, and Ethics of Care in the Cultural Commons." *Current Anthropology* 61 (S22): S289–S302.
OpenGlam. n.d. "A Global Network on Sharing Cultural Heritage." https://openglam.org.
Padilla, Thomas G. 2018. "Collections as Data: Implications for Enclosure." *College and Research Libraries News* 79 (6): 296.
Parry, Ross, ed. 2013. *Museums in a Digital Age*. New York, NY: Routledge.
Picheny, Michael, Zóltan Tüske, Brian Kingsbury, Kartik Audhkhasi, Xiaodong Cui, and George Saon. 2019. "Challenging the Boundaries of Speech Recognition: The MALACH Corpus." *arXiv preprint* arXiv: 1908.03455.
Pittock, Murray. 2018. "The Scottish Heritage Partnership Immersive Experiences: Policy Report." University of Glasgow. http://eprints.gla.ac.uk/198201/1/198201.pdf.
Ridge, Mia, ed. 2014. *Crowdsourcing our Cultural Heritage*. Farnham: Ashgate Publishing.
Rutner, Jennifer and Roger C. Schonfeld. 2012. "Supporting the Changing Research Practices of Historians." Ithaka S&R. https://sr.ithaka.org/wp-content/uploads/2015/08/supporting-the-changing-research-practices-of-historians.pdf.
Schön, Donald A. 1983. *The Reflective Practitioner: How Professionals Think in Action*. New York, NY: Basic Books.
Strauss, Anselm L. and Juliet M. Corbin. 1997. *Grounded Theory in Practice*. London: SAGE Publications.
Terras, Melissa. 2011. "The Rise of Digitization." In R. Rikowski (ed.), *Digitisation Perspectives*, 3–20. Rotterdam, Sense Publishers.
Terras, Melissa. 2016. "Crowdsourcing in the Digital Humanities." In *A New Companion to Digital Humanities*, edited by S. Schreibman, R. Siemens, and J. Unsworth, 420–39. Hoboken, NJ: Wiley-Blackwell. http://eu.wiley.com/WileyCDA/WileyTitle/productCd-1118680596.html.
Terras, Melissa, Stephen Coleman, Steven Drost, Chris Elsden, Ingi Helgason, Susan Lechelt, Nicola Osborne, Inge Panneels, Briana Pegado, Burkhard Schafer, Michael Smyth, Pip Thornton, and Chris Speed. 2021. "The Value of Mass-Digitised Cultural Heritage content in Creative Contexts." *Big Data and Society*. doi:10.1177/20539517211006165.
Thylstrup, Nanna Bonde. 2019. *The Politics of Mass Digitization*. Cambridge, MA: MIT Press.
University College London. 2021. "INST0044: Introduction to Digitisation." Information Studies. https://www.ucl.ac.uk/information-studies/inst0044-introduction-digitisation.
Warwick, Claire, Melissa Terras, Paul Huntington, and Nikoleta Pappa. 2008. "If You Build It Will They Come? The LAIRAH Study: Quantifying the Use of Online Resources in the Arts and Humanities through Statistical Analysis of User Log Data." *Literary and Linguistic Computing* 23 (1): 85–102.
Wingo, Rebecca S., Jason Heppler, and Paul Schadewald. 2020. *Digital Community Engagement: Partnering Communities with the Academy*. Cincinnati, OH: University of Cincinnati Press.
Zaagsma, Gerben. 2019. "Digital History and the Politics of Digitization." Digital Humanities, Utrecht University, July 8–12. https://dataverse.nl/dataset.xhtml?persistentId=doi:10.34894/USARQ6.
Zwitter, Andrej. 2014. "Big Data Ethics." *Big Data & Society* 1 (2). doi: 2053951714559253.