e-ISSN 2724-3923

magazén

International Journal
for Digital and Public Humanities

Vol. 1 – Num. 2
December 2020

Edizioni
Ca'Foscari

e-ISSN 2724-3923

magazén
International Journal for Digital and Public Humanities

Edizioni Ca' Foscari - Digital Publishing
Fondazione Università Ca' Foscari
Dorsoduro 3246, 30123 Venezia
URL https://edizionicafoscari.unive.it/en/edizioni/riviste/magazen/

magazén
International Journal
for Digital and Public Humanities

Semestral Journal

Head office Università Ca' Foscari Venezia | Venice Centre for Digital and Public Humanities | Dipartimento di Studi Umanistici | Palazzo Malcanton Marcorà | Dorsoduro 3484/D, 30123 Venezia, Italia | magazen@unive.it

Publisher Edizioni Ca' Foscari - Digital Publishing | Fondazione Università Ca' Foscari | Dorsoduro 3246, 30123 Venezia, Italia | ecf@unive.it

[ve]dph
Venice Centre for Digital and Public Humanities

PEER REVIEW OPEN ACCESS

Certificazione scientifica delle Opere pubblicate da Edizioni Ca' Foscari - Digital Publishing: tutti i saggi pubblicati hanno ottenuto il parere favorevole da parte di valutatori esperti della materia, attraverso un processo di revisione anonima sotto la responsabilità del Comitato scientifico della rivista. La valutazione è stata condotta in aderenza ai criteri scientifici ed editoriali di Edizioni Ca' Foscari.
Scientific certification of the works published by Edizioni Ca' Foscari - Digital Publishing: all essays published in this volume have received a favourable opinion by subject-matter experts, through an anonymous peer review process under the responsibility of the Scientific Committee of the journal. The evaluations were conducted in adherence to the scientific and editorial criteria established by Edizioni Ca' Foscari.

magazén
Vol. 1 – Num. 2 – December 2020
e-ISSN 2724-3923

Table of Contents

FUSIONS

URL https://edizionicafoscari.unive.it/en/edizioni/riviste/magazen/2020/2/
DOI http://doi.org/10.30687/mag/2724-3923/2020/02

Fusions

edited by Franz Fischer, Diego Mantoan, Barbara Tramelli

e-ISSN 2724-3923

magazén
Vol. 1 – Num. 2 – December 2020

By Way of Another Editorial on Fusions in the Digital and Public Humanities

Franz Fischer
Università Ca' Foscari Venezia, Italia

Diego Mantoan
Università Ca' Foscari Venezia, Italia

Barbara Tramelli
Università Ca' Foscari Venezia, Italia

Summary 1 Kickstarting an Open-Ended Debate at International Level. – 2 Five More Contributions to the Topic of 'Fusions'.

1 Kickstarting an Open-Ended Debate at International Level[1]

This second issue of *magazén* closes the inaugural volume 2020 and thematically follows high on the heels of the first one, as they are both connected to the topic of 'fusions', a term which in our intentions functions as a picklock to investigate recent developments in the wider field of digital and public humanities. Earlier this year, we invited scholars to ponder how this definition could be useful to interpret the field's attempt in the past few years to create a canon or, at least, a set of criteria for its own inception as an academic discipline. Through an international call for papers we thus wanted to open a platform that would allow theoretical debates, methodological reflec-

1 This introduction paper was mutually agreed on by the authors who acted as curators of *magazén*'s inaugural volume 2020, divided in two issues, with the help of the journal's editorial board.

tions, as well as the examination of particular case studies ranging from textual scholarship, history and art history to cultural heritage studies and archaeology. Our conviction is that digital and public humanities are still informed by ongoing mergers, interrelations, interpenetrations, interdependencies, and cross-contaminations that shape their very research processes and approaches. The response to our call for abstracts – which we issued twice, in March and then again in July 2020 – was astounding with over sixty proposals from all over the world and from scholars at different levels of their careers, spanning from PhD candidates to senior researchers and full professors. Given this high figure and the subsequent necessity to thoroughly peer review the selected articles we resolved to split the proposals in two yearly volumes, hence in 2021 we will pursue the topic of 'fusions' even further, though with a slightly different take. For the present volume we inaugurated a particularly strict selection process, in order to present our scholarly audience with papers of the best possible standard, such as to strengthen and legitimise this novel and interdisciplinary field towards the wider – and sometimes quite skeptical – domain of the humanities. For this purpose, our editorial board arranged a preliminary selection on the basis of the submitted abstracts, in order to ask for full papers only to those prospective authors that we deemed interesting for the aims of this year's topic. All submitted papers then underwent the scrutiny of a double blind peer review process with two expert scholars for each contribution. For the sake of statistics, out of approximately thirty proposals we considered upon our call for abstracts, eight were immediately rejected, while eighteen papers were eventually submitted by the scheduled deadline and only ten made it into the final volume. It was indeed a tough selection process that involved reviewers from major universities and research centres of international renown, thus it should speak for the high quality of the published papers that just a third of the initial proposals were eventually taken into consideration.

While this issue may close *magazén*'s first volume, the inaugural work for this journal is not yet over. As a matter of fact, considering the two adjectives our very research field comprises – digital and public – we are committed to develop a dedicated web version of the journal, which will allow for better categorisation and searchability of the final papers. Furthermore, we will attempt to exploit, for scholarly reasons, all possibilities offered by digital publishing that render a different presentation logic and reading experience, such as plain vertical scrolling, the insertion of various media and hyperlinks, as well as interrelated tagging. Hence, the provision of general categories and subdomains becomes paramount, not just to organise the papers of our journal, but rather to discern them from the perspective of a methodological structure for the entire field of digital and public humanities that conceptually focuses on contact points, similar-

ities, and interconnections among the various disciplines it is comprised of. As already mentioned in the introduction to the first issue of this volume, we thought of five relevant dimensions that form the basis of our categorisation system. Specifically they refer to: 1) the kind of materials observed or employed, which will be divided among monuments, documents, sound, movement, works of art or born digital artefacts; 2) the media of representation, since it may influence the content reception, be they image data, textual data, audio-visual data, 3D data or else; 3) the applied methods of research, which comprise modelling, epistemology, collection, processing, visualisation, analysis, hermeneutics; 4) the modes of sharing, thus focusing on the public aspect of the research, which may employ various forms of publication, participation, communication, preservation, or afterlife; and 5) actors, factors and agents of the chosen field, since they really constitute the analysed domain by determining its boundaries and behavioural patterns, hence one will highlight features referring to society, institutions, communities, technology, and the environment discourse. This structure consisting of five domains and their subcategories has an analytical purpose, of course, and may be revised in the future according to the advancement of the field itself or because of radical technological change, but so far we believe it is suitable in order to enshrine and subdivide the various aspects that hold relevance to our research domain. We hope future readers of our hypertext version will appreciate this set of transversal domains that we would like to address with due awareness, in order to contribute to establishing digital and public humanities as a coherent field inside academia. If we succeed in this task, even though only in part, it would be our greatest pleasure given that it is inscribed in our founding principles that we need to create a basis for the collaborative development of durable, reusable, shared resources for research and learning in the field of digital and public humanities. In a way, our aim is really just to set up an open, international, and interdisciplinary platform, as if we were to provide a nice venue and furnishing for our proverbial *magazén*, the public house where at the time of the Venetian Republic everyone was invited to share, talk, bargain, harangue, and exchange ideas, experiences and objects.

2 Five More Contributions to the Topic of ‘Fusions’

Following on the lines of the first issue, the authors chosen for this second venture start from a historical overview to unfold the interrelations and connected dynamics which occurred in various subdomains of the digital and public humanities. Hence, the first three contributions tackle the concept of ‘fusions’ analysing the gradual convergence of different fields at a methodological level, while the

e-ISSN 2724-3923

last two papers are dedicated to two intriguing case studies that examine the liminality between cultural production and consumption with regard to museum display in the digital environment.

To begin with the first contribution, Barbara Heinisch enquires on the meanings of what is now referred to as 'citizen humanities', that is, the public participation in scholarly research, and their connections to the digital and public humanities. Illustrated by a citizen linguistics project, she argues that the mutual influence on academic practices is reflected in novel ways of knowledge co-production, shared authority, and societal transformation.

In the second paper, Chris Beausang takes up the challenging task to retrace the history of computational literary criticism by dividing it into three main epochs, analysing the turning points and characterising each one in their connections to traditional literary criticism.

In the third contribution of this issue, Enrica Salvatori underlines the intrinsic and long-established relationship between historians and information technology, presenting current Italian projects in the digital humanities derived from national conferences of the Italian Association for Public History (AIPH) and the Association of Humanistic Informatics and Digital Culture (AIUCD).

Coming to the conclusive case studies of this issue, Trilce Navarrete and Elena Villaespesa analyse the online fruition on Wikipedia of more than one hundred paintings from the Metropolitan Museum of Art, arguing that digital cultural consumption can open and foster new ways of utilizing and experiencing art collections outside the traditional art context, and that this new form of digital consumption is reflected in the strategy which is increasingly adopted also by other important museums.

In the closing paper, Kathryn Simpson and Lois Burke present the understudied topic of 'children's writings', which always struggled to gain a place in traditional literature as well as in the museum space. They engage in a fruitful discussion on how to use digital tools to display the agency of these writings, convincingly arguing that presenting children's collections digitally can be useful in order to create a space for experimentation and exchange between institutions, objects and visitors.

Finally, we must again express our most sincere acknowledgment to the many scholars involved in this venture in various roles and with different capacities: the numerous colleagues and friends at the Department of Humanities engaged in the establishment of our research centre; the external experts of our advisory board for their trust and consideration; all contributors, peer reviewers and members of the editorial board for their strong commitment and smooth collaboration; the team of our publisher, Edizioni Ca' Foscari, directed by Massimiliano Vianello, for the tireless job necessary to deliver both issues of volume 2020 with perfect timing. Let us hope 2021

will bring about better times for all of us, though we will certainly keep on working on another volume of *magazén* in a field of growing relevance for the humanities, particularly in a period of social distancing and limited movement.

e-ISSN 2724-3923

magazén
Vol. 1 – Num. 2 – December 2020

Citizen Humanities as a Fusion of Digital and Public Humanities?

Barbara Heinisch
Universität Wien, Österreich

Abstract Digital and public humanities have gained a foothold in academia, but very little is known about citizen humanities, which is referring to the engagement of the general public in scholarly research. Although the term is new, public participation in the humanities, either as the citizens' contribution of intellectual effort or knowledge to academic research, or as the contribution of resources and tools, looks back on a long tradition. The citizen humanities range from the creation of dictionaries, the transcription and annotation of historical records to the decoding of ancient Egyptian papyri. While the digital humanities provide the citizen humanities with data, tools and techniques, the public humanities offer the means of engaging diverse publics in research activities. After embedding the citizen humanities theoretically in the responsible research and innovation paradigm, this paper will illustrate how digitisation and public involvement laid the foundations for today's citizen humanities. With a focus on the fusion of digital and public humanities in citizen humanities, this paper will demonstrate the mutual influence on practices (of research). This influence is not only reflected in the approaches to research, analysis, communication, and dissemination but also in the citizen humanities' novel ways of knowledge co-production.

Keywords Responsible research and innovation. Third mission. Crowdsourcing. Public participation in research. Participatory research. Citizen science. Public engagement.

Edizioni Ca'Foscari

Peer review

Submitted	2020-07-17
Accepted	2020-10-27
Published	2020-12-22

Open access

Citation Heinisch, B. (2020). "Citizen Humanities as a Fusion of Digital and Public Humanities?". *magazén*, 1(2), 143-180.

DOI 10.30687/mag/2724-3923/2020/02/001

1 Introduction[1]

The humanities are subject to continuous change. As diverse as the broad spectrum of disciplines it encompasses, ranging from philosophy, history, archaeology to literature and ancient and modern language, are the topics under investigation and the methods and technologies applied. Under the umbrella of the humanities, different forms to study human culture have emerged, such as the digital humanities or the public humanities.

While the digital and public humanities have gained a foothold in academia, very little is known about citizen humanities, which is referring to the engagement of members of the public in scholarly research. Although the term is new, the concept of public participation in the humanities and in activities of cultural heritage institutions is old. Members of the public have contributed their resources, effort and knowledge to academic research or initiated research themselves also in the past. Participants in projects of the citizen humanities or of cultural heritage institutions contribute to the creation of dictionaries, the transcription and annotation of historical records or the decoding of ancient Egyptian papyri.

Both the digital humanities and the public humanities contributed to the development of the citizen humanities. Digitalisation and public involvement laid the foundations for today's citizen humanities, impacting processes, approaches, and practices of research in this wider field being considered a fusion of digital and public humanities. While the digital humanities provide the citizen humanities with data, tools, techniques and infrastructures that do not only facilitate humanistic inquiry but also communication and collaboration with different actors, the public humanities offer the means of communication and ways of engaging diverse publics in research activities.

The citizen humanities are not only influenced by but do also exert impact on the digital and public humanities. This mutual influence is characterised by the materials collected or analysed, the methods applied, the media of (knowledge and data) representation and the ways of collaboration (between researchers and citizens). The citizen humanities thus can lead to mutual exchange and knowledge co-production.

In order to identify the contributions of the digital and public humanities to the citizen humanities and to identify the aspects that shaped the fusion of digital and public humanities in citizen humanities, the third mission paradigm, the pillars of responsible research and innovation and a citizen linguistics case study are used as the basis of analysis. This work will generate fresh insight into the com-

1 This research was partially supported by the Austrian Science Fund (FWF): TCS 57G. Thanks also go to the anonymous reviewers.

monalities of the public humanities and the digital humanities with the citizen humanities and their interrelations with responsible research and innovation. Nevertheless, this study is unable to encompass the entire sphere of public humanities, digital humanities and the long tradition of collaboration between academia and members of the public, ranging from different technologies and methods to different sub-disciplines, such as public history or public archaeology.

This paper first gives a brief overview of the emergence of the citizen humanities before aligning the core aspects of the public humanities (PH) and digital humanities (DH) with the citizen humanities.

The term 'citizen humanities' has come to be used to refer to "citizen 'science' in the humanities" (Heinisch et al. forthcoming). Since it is derived from the notion of 'citizen science', it is important to shed some light on citizen science before proceeding to elaborate on the citizen humanities.

1.1 Citizen Science

Several definitions of citizen science (CS) have evolved. According to the *White Paper on Citizen Science for Europe*, citizen science is basically understood to mean the "general public engagement in scientific research activities when citizens actively contribute to science either with their intellectual effort or surrounding knowledge or with their tools and resources" (Serrano Sanz et al. 2014, 8). Interestingly, "public engagement" is mentioned in the *White Paper*'s definition, which is also one of the pillars of responsible research and innovation, which will be discussed later.

In other words, CS can also be described as science done by people (Silvertown 2009, 470), i.e. non-professional academics engage in scientific investigations and ask questions, collect or analyse data or interpret results (Miller-Rushing et al. 2020, 17). This means that being a volunteer in a medical trial or responding to a social science survey do not qualify as CS (Haklay 2013, 2).

CS ranges from "large-scale data collection" to "engaging public perspectives and knowledges in science discourse and policy making" (Shirk et al. 2012, 26). Thus, the understanding of CS is rather broad, ranging from crowdsourcing to participatory action research.

While the term 'citizen science' is rather new, the underlying concept is old. Members of the public without professional training in the field of research have been contributing to science for centuries. Either members of the public carried out academic investigations independently or they collaborated with academic experts (Reiheld, Gay 2019). For centuries, citizens have made observations and records, for example of flora and fauna, phenology, weather or astronomy (Miller-Rushing, Primack, Bonney 2012, 285).

The emergence of CS was shaped by two strands of CS: Democratisation of science (Irwin 1995) and public participation in scientific research (Bonney et al. 2009). The first strand addresses the relationships between citizens and science and the responsibility of science towards society. Thus, it has a clear relation to responsible research and innovation and service to society based on two assumptions: "a science which assists the needs and concerns of citizens" and "a form of science developed and enacted by citizens themselves" (Irwin 1995, xi). The second strand subsumes different models of public participation in scientific research under three categories according to the degree of public involvement and the control participants can exercise in different steps of the research process (Bonney et al. 2009, 11).

CS has recently been fuelled by technological developments, entailing new means of communication, collaboration and data. Scholars ask the public to support them in their research, beyond being the mere subject of investigation.

Different reasons for the current popularity of CS have been proposed. First, democratised knowledge production may not only lead to societal transformation but also to academic breakthroughs (Bela et al. 2016, 990). Another explanation are social movements, such as the environmental justice movement or the women's health movement, that call for social change and intervene in science, technology or medicine to make them more participatory and inclusive. For this purpose, they use and contest scientific expertise and demonstrate the value of local and indigenous knowledges. This way, academic research is subjected to increased public scrutiny, opened up to participation and different views of knowledge, thus, paving the way for the acceptance of citizen science. Simultaneously, science is undergoing neoliberal transformations regarding funding and organisation that lead to a decline in public funding and, thus, to an increased interest in using citizen science to conduct research with the help of volunteers. Other factors are a society oriented towards risk management that requires continuous monitoring of the environment, and the scientisation of politics (Kimura, Kinchy 2016, 335-7).

1.1.1 Citizen Humanities

The European Citizen Science Association (ECSA) sees the citizen humanities as inherent part of citizen science. However, the fact that "science" primarily comprises "natural sciences" and that citizen science has a strong focus on studies of the environment and biodiversity (Tweddle et al. 2012, 1) resulted in new strands entitled 'citizen social science' and 'citizen humanities'. The major difference between these three strands is the object of investigation. While citizen science encompasses natural science disciplines, such as biolo-

gy, chemistry and physics, citizen social science studies societies and the citizen humanities cover literature, language, philosophy or history (Heinisch et al. forthcoming). Additionally, research approaches and schools of thought may differ significantly.

In the humanities, public engagement can take various forms, including the transcription of handwriting, tagging of text or images, entry of structured data, participation in discussions, commenting or doing oral history and recording personal memories and experiences (Hedges, Dunn 2018, 1) as well as correcting content, cataloguing, contextualising, mapping, georeferencing or translating content (Dunn, Hedges 2012, 21).

1.1.2 Related approaches

As diverse as the CS landscape are the designations used for the different types of participatory research practices or engagement of non-academics in scholarly research. Related terms that are sometimes used interchangeably are community research, community science, crowdsourced science, civic science, amateur research, public participation in science, (academic) crowdsourcing, (community-based) participatory research, participatory science, participatory action research etc. (Pettibone, Vohland, Ziegler 2017; Kullenberg, Kasperowski 2016, 2). Since they originate from different schools of thought and, thus, emphasise different aspects, and in some cases even different degrees of public engagement in research, they cannot be considered synonyms.

For example, participatory research, which is often associated with the social sciences, puts the participating humans, including their perspectives, learning processes and their empowerment at the centre. It is not a purely academic endeavour but always a joint project with non-academic, societal actors who are considered co-researchers. The characteristics of participatory research are to conduct research on and influence social reality. Participation refers to both participation in research and participation in society with paying particular attention to the actors' empowerment and values that guide the research endeavour (Unger 2014, 1-2).

Another differentiation can be made between community-initiated projects (bottom-up CS activities) and researcher-initiated projects (top-down activities, where a researcher has a clear hypothesis or research assumption and already defined the research process who needs participants to contribute to smaller tasks and activities in the research project, e.g. collecting or analysing large amounts of data).

The latter form of the citizen humanities is also referred to as '(academic) crowdsourcing' in the humanities (Hedges, Dunn 2018). "Crowdsourcing is the process of leveraging public participation in

or contributions to projects and activities" (1). Usually, an organisation or researcher calls for assistance from volunteers who undertake small portions of a task to solve a problem in humanities research. Although crowdsourcing sometimes carries a negative connotation, in the humanities and cultural heritage institutions, it is rather considered a contribution towards a shared and significant goal of community and intellectual value (Terras 2016, 427; Tanner 2015).

Large-scale crowdsourcing projects in the culture and heritage sector include digitisation projects, such as the correction of optical character recognition errors in digitised material, the transcription of historical records or playing games to improve the metadata of collections (Terras 2016, 424). Since the (digital) humanities are intertwined with culture and (cultural) heritage, it is sometimes not easy to draw boundaries between academic crowdsourcing in the humanities and crowdsourcing in and for cultural heritage institutions. Even if crowdsourcing projects in cultural heritage institutions have a major focus on sorting, labelling or formatting historical data, these data can provide the raw materials for academic research aimed at inquiring human culture (426).

Thus, crowdsourcing adopts a top-down approach, in which researchers specify what they need from the crowd and the crowd contributes small pieces to a project. The citizen humanities have a wider scope than crowdsourcing. Although the citizen humanities also include crowdsourcing, such as tagging, transcribing or annotating research data, they cover a broader range of activities and would also encompass forms of participatory (action) research or co-creation, such as initiatives in which the community has the lead or shares stronger responsibility with academics or co-develops research questions, research designs or project management.

1.1.3 Long tradition

Participatory (research) practices and public engagement as well as collaborative approaches in the humanities look back on a long tradition. Related movements that (also) rely, more or less, on the collaboration between research institutions and the public are, among others, democratic education, settlement houses, service learning or community development. While service learning states that higher education institutions are responsible for helping their students develop skills that are required for being an active citizen, including finding solutions to (public) problems that are solved through collective labours (Boyte, Farr 1997, 7), community development "involves actions based on values and principles" (Kenny, MacGrath 2018, xx) that address issues impacting humans and their conditions while starting with the communities' perspectives, thereby aiming at the

'what should be'" (xxi). The 'new scholarship' is defined as a "scholarship of *action*, a scholarship of *practice* that takes place both *in* and *with* the community" (Fitzgerald, Primavera 2013, 131; emphasis in the original) which challenges the epistemologies of universities and the scope of legitimate knowledge. It offers a more socially utilitarian and more egalitarian model on the definition, acquisition, communication, use, and evaluation of knowledge, in which academia is only one among many proprietors and distributors of knowledge. The new scholarship includes a collaboration between academia and the community who share their knowledge and resources to tackle pressing societal problems and resulting in rich and deep relationships (131-2).

In comparison to today's citizen humanities, these participatory practices have a stronger focus on (co-)production of knowledge for the purpose of solving public or societal problems, including inequalities. To succeed, these approaches require participation, engagement, collaboration and partnerships that rely on values, such as trust, inclusion and transparency. This interest in participatory approaches to research creation and cross-disciplinary alliances is symptomatic for the public turn at universities. Various forms of participatory practices (that can also be found under the heading of 'going public') are committed to collaboration at all project stages to challenge power structures and increase a project's impact. Any collaborative endeavour requires time and resources to build relationships and trust, clarify expectations and include reflection. Moreover, researchers are required to be open, i.e. open to co-create, to new methods, new tools, new ideas, media and relationships, especially if cultural and disciplinary boundaries are crossed. Moreover, ideas get better if they are discussed by different people. Researchers have to work outside their comfort zones and engage in creative ways and try new approaches. Nevertheless, this in contrast to the boundaries of academic disciplines, with a closed set of methods and a level of authority. Cross-disciplinary methodologies include photo-voice, community mapping, digital storytelling or participatory archiving. Moreover, this public turn resulted in a rethinking of research creation, including how, why, with whom and for whom research is done. Funding bodies accelerated this development by providing grants for creating new spaces of (sustained) conversation and co-creation between university researchers, (artists) and the wider community. Not only artists and activists, but also feminists and indigenous researchers "contributed to a deeper reflexivity about the situated nature of research, demanding transparency and raising critical questions about who owns and benefits from any research endeavour" (Miller, Little, High 2017, 4-6). The notion of 'impact' as used in today's research jargon may be less important than the relationships fostered (The CRESC Encounters Collaborative 2013, 2). This shows that today's citizen humanities are preceded by different movements

e-ISSN 2724-3923

and approaches of interaction between universities and the public that may also influence the way the citizen humanities are conducted.

1.2 Digital Humanities

Several definitions of digital humanities (DH) have been proposed. DH "involves the application of computers and various digital tools and resources to the study of Humanities" (Siemens 2010, 39). The DH are the overlap between traditional humanities and computational methods and digital tools (Burdick et al. 2012, 3). They are characterised by scholarly methods in the form of computer-based methods that support the creation, analysis and dissemination of research and teaching in the humanities (Hughes, Constantopoulos, Dallas 2016, 152). Thus, the DH designate "humanities research in the digital era" and "digital scholarship in the humanities" (Vanhoutte 2013, 144) characterised by three aspects: data, tools and collaboration. Technological advances lead to new digital research methods and tools for data analysis (including natural language processing, artificial intelligence and machine learning), the new availability of (large amounts) of data sources due to digitised collections, born-digital content or open data; as well as collaborations that encompass interdisciplinary, intersectoral and international collaboration that inject new ideas and perspectives into projects (Hedges et al. 2019, 7-8). The promises of the DH lie in the sharing of data, results and tools to distribute knowledge in a fair and broader manner and in new ways of representing, sharing and discussing knowledge (Sample 2013, 255-7).

The "Digital Humanities Foresight" study identified five major topics that should be the foundation for a DH research agenda. In addition to research infrastructures, the establishment of the digital commons, artificial intelligence and evaluation and impact metrics and methods, this study also put an emphasis on public engagement (Hedges et al. 2019, 11), which is at the core of the citizen humanities.

1.3 Public Humanities

In general, the academic humanities are targeted at academic professional audiences, while the public humanities (PH) are "oriented to nonspecialist audiences and nonacademic careers" (Ellison 2013, 289). The humanities are usually conducted within academic and institutional frames and public refers to something outside of these institutions (Carton 2009, 11-12). While the term PH is often used for non-academic humanities careers (Ellison 2013, 291), other authors regard the PH as "acts of reflexive, collaborative meaning-making informed by a collective good" (Cox, Tilton 2019, 129-30). "Public

humanities is about finding both practical and conceptual locations, spaces, and translations between the various kinds of humanities work that people are doing privately, publicly, in groups, in families, in religious communities – as well as in universities" (Carton 2009, 12). In either case, partnerships are at the heart of the PH. PH encompass collaborative research projects with communities, public or online lectures, conferences planned with regional partners, (social) media coverage or exhibitions (Ellison 2013, 293; Wickman, Browne 2014).

Thus, the PH are strongly related to outreach science communication and public engagement. Usually, the PH engage members of the public to participate in conversations and reflect about topics and ideas. Nevertheless, public history and public archaeology, for example, usually have a stronger focus on the work done by the public, i.e. public history "as a mixture of history for the public, about the public, and by the public" (Cole 1994, 12).

1.4 Digital Public Humanities

As the terms already suggest, the DH are characterised by the digital, while the PH are defined by the public, sometimes also understood as the responsibility of serving communities (Brennan 2016, 384). "Public digital humanities, then, should be identified by the ways that it engages with communities outside of the academy as a means for doing digital humanities scholarship" (384). However, only being present on the web does not qualify as digital public humanities (DPH).

The DPH invite non-academic audiences to contribute to scholarly research. Since the audiences are contributors and users alike, the digital technologies should be subject to user-centred design, including functionalities, languages, navigation etc. in order to make them feel welcome.

What the DPH share with the citizen humanities are that the participants or the public are involved in the project as early as possible and not only at the end of the project as part of outreach activities to disseminate findings. Another commonality between the DPH and the citizen humanities are the relevance for the community. Although the DPH may address shared responsibility, this may be less prominent in the citizen humanities, for example in academic crowdsourcing in the humanities. Here, often scholars have the lead, make the decisions and assume responsibility for the project.

While Brennan (2016) rather describes the joint design of digital technologies for joint research agendas as core aspects of the DPH, other authors attribute a transformative character to the DPH shaped by co-creation, shared authority and collaboration that should ensure

unpredictable results, a shared mission and collaborative meaning-making (Cox, Tilton 2019, 130-1). The latter comes close to the understanding of the citizen humanities presented in this paper.

However, the roots are different since the DPH are strongly related to public history, which was shaped by volunteers preserving community objects or stories. Moreover, the PH sometimes have a service character (Brennan 2016, 385) that is rather not at the core of the citizen humanities.

Another concept not addressed in this paper is the "engaged humanities" that raise related issues, such as community engagement versus the political economy of higher education, institutional barriers to engagement and public scholarship, putting the discipline or the community first, educating students and practitioners, the necessity for asset mapping of community and participants, turning projects into partnerships, reexamining course goals, learning outcomes and assessment, institutionalising engaged courses, balancing workloads for faculty, students and community partners as well as diversity and engagement (Jay 2012, 57-60).

1.5 Third Mission

CS is strongly linked to the third mission of universities. In addition to research and teaching, universities are required to exert impact (on society) beyond academia (Henke, Pasternack, Schmid 2018, 57). These third mission (or third stream) activities refer to "the generation, use, application and exploitation of knowledge and other university capabilities outside academic environments" (Molas-Gallart et al. 2002, iii–iv). They cover various types of interactions between a university and society, between academia and non-academic practice. The main target audience of these (communication) activities are non-academic communities, whereas engagement plays a central role (Molas-Gallart et al. 2002, 5). The term third mission is generally understood to cover three areas: continuing education, research and knowledge transfer, and societal commitment. Examples of societal commitment include civic engagement (creating a democratic citizenship), community outreach (giving knowledge to society), community service (integration of university members in social projects with mutual learning successes), service learning (societal commitment as part of the curriculum and preparation of students for societal projects), social entrepreneurship (changing societal conditions through entrepreneurial behaviour) and widening participation (broadening the target group of higher education). This shows that the third mission encompasses more than knowledge and technology transfer. It considers universities as part of society that (voluntarily) contribute to developments in society. Therefore, third mission activities are

characterised by interactions with actors outside academia and by contributing to societal development interests beyond research and teaching (Henke, Pasternack, Schmid 2018, 45-80).

Another concept often used in connection with third mission and CS is responsible research and innovation.

1.6 Responsible Research and Innovation

Responsible Research and Innovation (RRI) is defined as "the on-going process of aligning research and innovation to the values, needs and expectations of society" (Rome Declaration on Responsible Research and Innovation in Europe 2014, 1). Therefore, multiple stakeholders, including researchers, policy makers, business, NGOs and civil society assume responsibility and must be part of the research and innovation process. Stakeholders should be engaged from the very beginning to improve creativity and enhance academic quality throughout the process. Furthermore, RRI is aimed at achieving acceptable, desirable and sustainable innovation. Therefore, excellence is more than innovative discoveries and covers responsibility, openness and the co-production of knowledge (Rome Declaration on Responsible Research and Innovation in Europe 2014, 1) to align research outcomes and processes with the values, expectations and needs of society. One major driver of RRI is the European Commission's Horizon 2020 framework programme in which projects should tackle the grand challenges, such as climate change, energy or health to foster intelligent, sustainable and overall growth in Europe (enRRICH 2016). These grand societal challenges need to be tackled by engaging all societal actors in jointly finding solutions. The major characteristics of grand societal challenges are their complexity, their mutual dependency and intertwined social, economic and ecological issues on local and global levels. Their multidimensional, transdisciplinary and systematic nature requires new approaches and perspectives to allow for complex transformation processes (Lindner et al. 2016, 41). This also means that societal needs are addressed by participatory approaches that engage all actors throughout the entire research and innovation process (EU 2012). This should ensure that new perspectives and otherwise unnoticed solutions (and also risks) come to the fore when addressing societal challenges so that sustainable and inclusive solutions build trust between all institutions and citizens involved in research and innovation (Rome Declaration on Responsible Research and Innovation in Europe 2014, 1).

RRI considers academia, economy and society as a whole and has several dimensions and aims. Moreover, RRI does not only focus on the final product or outcome of the research and innovation process but also on the process itself (Schomberg 2012, 50).

RRI helps researchers challenge own underlying assumptions on an individual and institutional level (reflexivity), involve the population in academia, research and innovation processes (participation, inclusion, stakeholder engagement), consider the perspectives and needs of society in research and innovation processes and adapt the development accordingly (responsivity), anticipate developments and achieve socially desirable aims (anticipation), promote equal opportunities, gender equality and diversity, consider ethics in research and innovation and strengthen the researchers' integrity, increase access to research (open access) and improve science education (RRI-Plattform Österreich 2020).

Therefore, the outcomes of RRI are threefold: learning outcomes that result in responsible and empowered actors; sustainable, socially desirable and ethically acceptable outcomes of research and innovation and solutions to societal challenges (RRI Tools 2020).

The six pillars of RRI are ethics, gender equality, open access and data, science education, public engagement and governance, thus covering various areas of the relationship between research and innovation, on the one hand, and society, on the other (RRI Tools 2020).

Having defined what is meant by responsible research and innovation and having discussed the relationship between RRI and CS, the following section will explain the RRI pillars before proceeding to examine the contributions of the PH and the DH to the citizen humanities according to the RRI pillars.

2 Analysis

The objective of the analysis is to identify the commonalities of the DH and PH with the citizen humanities and testing the assumption that the citizen humanities are a fusion of the DH and PH. To compare and find similarities between these three strands of the humanities, the six RRI pillars were used as a common basis.

"RRI entails engaging all actors (from individual researchers and innovators to institutions and governments) through inclusive, participatory methodologies in all stages of R&I processes and in all levels of R&I governance (from agenda setting, to design, implementation, and evaluation)" (RRI Tools 2020). Therefore, the concept of RRI gives researchers the opportunity to seize the population's creativity and knowledge either through open innovation or CS (RRI-Plattform Österreich 2020, 3).

This shows that CS and RRI are intertwined. The opinion "Toward an International Definition of Citizen Science" specified inclusion and exclusion criteria for projects and initiatives that qualify as CS. The major categories addressed in this opinion are scientific standards (based on scientific questions or hypothesis testing, the methods and

the rationale for generating new knowledge or new methods), collaboration (e.g. active involvement of participants in the research process or the added value for all persons in the project), open science, communication (transparency, dialogue among interest groups via various communication channels, data quality), ethics (adherence to ethical standards, data policy, governance, informed consent and inclusiveness) and finally, data management (including a data management plan) (Heigl et al. 2019, 8091).

ECSA's ten principles of CS, for example, also address RRI aspects, such as ethics, governance, public engagement and open access and data etc. (European Citizen Science Association 2015, 1). This is also found in the "Quality Criteria for Citizen Science Projects on Österreich forscht", the Austrian citizen science platform. These quality criteria cover RRI pillars such as ethics, governance (covered by the criteria collaboration and communication) or open science (Heigl et al. 2018).

Additionally, since there are different degrees of public participation in science (Haklay 2013; Bonney et al. 2009), the extent to which the RRI pillars of governance and public engagement are covered in CS projects also depends on the degree of public participation. Co-created approaches (Bonney et al. 2009) or extreme CS (Haklay 2013) consider governance and public engagement to a greater extent. Governance in co-created CS projects means shared responsibility and provision of related instruments; and public engagement means working with societal actors through the entire research process (and not only during data collection or analysis) to align research processes with the needs, values and expectations of society. The RRI pillars of governance and public engagement require that various stakeholders must be involved in the research process. However, not all CS projects are able to recruit multiple stakeholders.

When evaluating CS projects, RRI aspects and third mission also play a role. Evaluation frameworks may consider "three dimensions of participatory science: (i) scientific impact, (ii) learning and empowerment of participants and (iii) impact for wider society" (Kieslinger et al. 2018, 81). CS projects may also be evaluated based on their scientific output, citizen participation and involvement in research processes, education (scientific literacy or disciplinary knowledge), nourishing a new consciousness of socially relevant topics, transparency regarding roles, functions and use of the outcomes as well as data protection and privacy and long-term consequences and sustainability (Pettibone et al. 2016, 21).

In the following, the six RRI pillars are used to analyse the commonalities between the digital, public and citizen humanities and address the fusion of the DH and PH in the citizen humanities. Since an in-depth analysis is beyond the scope of this paper, two aspects are studied in more detail: ethics as a core discipline of the humanities

and public engagement, which is at the centre of the citizen humanities. Additionally, a citizen linguistics project serves to illustrate the implementation of RRI in the citizen humanities.

2.1 Ethics

The RRI pillar of ethics is based on the shared values in European society. When responding to societal challenges, not only legal aspects but also ethical standards must be observed. This also enhances the acceptability and relevance of research and innovation in society. Although ethics is often perceived as impeding research and innovation, the European Union sees ethics that is considered throughout the research and innovation process as fundamental aspect of research excellence in all domains, including the humanities (EU 2012), which are a branch of knowledge significantly contributing to the discourse on (research) ethics.

Ethics covers research integrity and good research practice, research ethics for the protection of objects of research and societal relevance and ethical acceptability of research and innovation outcomes (European Commission 2015, 6-7).

Research integrity and good research practice are at the heart of any academic research, including the DH and PH. These include the compliance with legal regulations, such as intellectual property rights and data protection regulations. These also apply to the citizen humanities, but additionally they explicitly focus on the ethics concerning the participation of certain groups of people as participants and their role and right to information. However, this is not related to the aspect of protection of objects of research but to societal relevance and ethical acceptability of research and innovation since members of the public are not the objects of research but perform scholarly tasks themselves.

Related to ethics in CS are also trust and the relationships between different actors in research, such as individuals and organisations. While crowdsourcing in the commercial sector is sometimes described as exploitation (of free and volunteer labour), in the cultural heritage sector and the humanities, crowdsourcing is justified with a long tradition of altruistic participation and volunteering in academic projects or cultural heritage institutions, such as libraries, museums, galleries or archives. Moreover, crowdsourcing means working for the social good in a variety of interesting fields. This includes, for example, the correction of OCR (optical character recognition) text of newspaper articles or the transcription of (handwritten) observation cards in museums. These corrections and transcriptions enabled (digital) access to sources and further research (Terras 2016, 427-8).

As mentioned before, CS encompasses a wide range of forms and projects. Especially from the viewpoint of democratisation of science (Irwin 1995), CS is envisaged as public engagement characterised by mutual dialogue on eye level and giving all persons an equal voice, thus decreasing the divide between science and the public (Riesch, Potter 2014, 109). To overcome this divide, transparency, trust and different forms of benefits are needed. However, traditional forms of involving members of the public may reinforce hierarchies that hamper humanists in reaching the impact they actually seek. Moreover, 'going public' and allowing for transparency alone do not make citizen humanities projects significant (Wickman 2016, 9). It may be hard to work in a non-hierarchical and reciprocal way and to establish trust, for example, in the case of commissioned research. Therefore, "the symbolic and material act of listening to participants' viewpoints" (The CRESC Encounters Collaborative 2013, 5) is crucial to establish an environment of trust.

The *ECSA's Characteristics of Citizen Science* mention ethics and explicitly require transparency throughout research, consensual involvement of the participants, research integrity and quality (ECSA 2020, 3). Transparency includes information about the topic, rationale and methods of the research project, the team and organisation behind the research, the research process, the use of (personal) data, the participants' contributions and the outcomes. Transparency also necessitates a permanent feedback loop with the participants and progress reports.

The benefits for participants can take many forms, including status, personal development, such as receiving training, gaining experience or gratification (Dunn, Hedges 2012, 16) as well as personal rewards, such as contributing to a good cause. Moreover, gamification, competition and other rewards can increase the motivation of participants. These encompass being part of a community or giving back to a community, achieving group goals or discovering new fields as well as competition or rewards in the form of rankings or badges for high achievers (Terras 2016, 426-7). Contributors are usually drawn to a project based on their passion for the subject. Additionally, they are also part of a community, which can develop dynamically and also develop mechanisms for mutual support and self-correction (Dunn, Hedges 2012, 2).

A major concern is the sustainability of the participants' contributions since projects may be short-term and data can be easily lost if there are no institutional resources for long-term storage and maintenance. Moreover, short-term projects also have the risk of petering out and frustrating participants if they are no longer continued although the community would be willing to further contribute.

The quality of the participants' contributions and the question if people without academic training can produce reliable academic

data are major concerns. To gather high-quality data, clear instructions and training are necessary. Scholars may also worry that participants may distort the dataset. To enhance the quality of the data, CS projects usually have mechanisms, such as validation of content by several other users or moderation or verification by experts that ensure that the quality meets a specified benchmark (Terras 2016, 427-8). However, even if there are well-designed software, useful manuals, data entry rules and various trained user groups, the material on which the participants work may be problematic, unrecognisable and include additional notations that question the previous interpretation of data (The CRESC Encounters Collaborative 2013, 9).

The quality of the results, e.g. of transcriptions or annotations, plays an important role to ensure that the research outcomes can be used for further humanistic inquires in the future. Here, sustainability comes into play in order to allow for the re-use of the project results for different purposes in the long term, including research and innovation. This includes the format in which the data are stored, interoperability and standards to ensure the re-use, re-purposing or integration into other projects (Terras 2016, 431-2).

Here, the DH offer various techniques to ensure both the quality and the sustainability of research data. These include the Text Encoding Initiative (TEI) guidelines that provide a framework for modelling, analysing and presenting textual data. However, the compliance with TEI varies between (citizen) humanities projects. Nevertheless, projects that make use of TEI for transcriptions have shown that users can follow text encoding guidelines. If the participants are trained according to (scholarly and pedagogical) standards, they can also broaden their competences (Terras 2016, 432). Terras emphasises that the role of the DH in the area of crowdsourcing is twofold (436). On the one hand, they can increase the understanding of and seize the opportunities offered by the method of crowdsourcing in the humanities. These opportunities may include outreach and public engagement and demonstrating the value of the humanities. On the other hand, they can give advice to crowdsourcing projects in the humanities or the areas of culture and heritage to create datasets that meet academic standards, are useful and are re-usable in academic research to promote the understanding of culture and history.

Despite the promises of new technologies and new data that allow researchers to pose new research questions, the (digital) humanities also have the responsibility to take a critical stance towards the application of digital methods and the types of data made available (Hedges et al. 2019, 13). The humanities should not only apply trendy technologies but put these technologies and related developments under scrutiny since ethics and values are at the heart of the humanities.

The humanities should critically reflect on and bring ethical considerations to the use of artificial intelligence as part of and as means

for research and innovation (Hedges et al. 2019, 13-14). This also applies to the DH, where scholars should put the impact of the digital on scholarly practice under scrutiny, in addition to using digital techniques to address research questions in the humanities. Moreover, the humanities are asked to reveal potential fields of tension that may arise in the citizen humanities.

The citizen humanities themselves are a field of tension since work that has been previously accomplished by professional scholars is now (partly) done by persons with not professional training in the field of research. This raises issues of data quality, professionalism, free labour and reliability. Moreover, it raises issues of trust between researchers and participants and trust in the results. Moreover, funding for CS could have also been used to fund professional researchers instead (Terras 2016, 431). The citizen humanities also challenge existing divides, academic power and thus, the understanding of who is entitled to produce knowledge and conduct research, i.e. scholars, and who should rather receive knowledge, i.e. the 'public'. The citizen humanities mean a shift from seeing members of the public only as an audience to considering them as active participants in framing and conducting research (Belknap 2015). At the interface between science, society and policy, CS also has a social mandate (Serrano Sanz et al. 2014, 18).

2.2 Gender Equality

In Horizon 2020, three aims address gender equality. First, research teams should promote gender balance to increase the number of women participating. Second, gender balance has to be ensured in decision-making. These two aims target the removal of barriers and combating discrimination of women in academic careers and decision-making processes. This should lead to (long-term) institutional change, such as structures that affect women's career progression in institutions, promotion of gender equality and reduction of (the unconscious) gender bias as well as adjustments to workplace arrangements. Third, the gender dimension has to be integrated in research and innovation content in order to enhance the academic quality and societal relevance of the outcomes since the behaviours, needs and attitudes of both genders are considered (European Commission 2015, 6; 2020c).

Gender equality is addressed by the citizen, public and digital humanities to a different extent and on different levels. In general, several funding schemes or organisational policies require a balanced gender ratio in research projects and may also require the inclusion of the gender dimension in research. Citizen humanities project have to address an additional level. In addition to the gender-balanced re-

search team as well as decision-making bodies, and the inclusion of the gender dimension in research, they may also aim at achieving a gender balance among the group of participants. However, since the participants are usually self-selecting, this can be hard to reach.

Nevertheless, inclusion also plays a significant role in CS, which is exemplified by various publications and initiatives that aim at addressing groups that have not participated in CS projects so far, including ECSA's Equity, Inclusion and Empowerment working group.

2.3 Open Access

Responsible research and innovation require transparent and accessible research and innovation. Open access "means giving free online access to the results of publicly-funded research (publications and data)" (EU 2012). The availability of research findings free of charge fosters knowledge circulation and the uptake of academic results by different societal actors, who, in turn, can boost innovation. Moreover, it can improve and make research more efficient. However, it also raises issues of intellectual property rights and the necessary infrastructure as well as collaboration among and dialogue with all societal actors in the research and innovation process (European Commission 2020b).

Open access covers the accessibility and ownership of scholarly information. Open access is key in the citizen humanities since the results of the research to which citizens contributed should be published in a way so that participants can freely access all project results that have been achieved with their help.

Despite the benefits open access entails, scholars also face different obstacles in this area. Especially intellectual property rights (also with regard to citizen humanities) and data protection regulations, such as the General Data Protection Regulation (GDPR) are identified obstacles to sharing data and findings (in the DH). Other barriers include funding and the digital divide (Hedges et al. 2019, 8). In some cases, researchers have to pay to publish their publications as open access publications.

Open access has the promise of being able to re-use data. Therefore, the FAIR (findable, accessible, interoperable and re-usable data) principles have to be applied to research data. Nevertheless, data sets may be dispersed, may have no uniform metadata or annotations, or do not comply with standards which makes them either hard to find or to re-use, including combining data from various sources.

A major contribution by the DH to both the citizen humanities and the PH are research infrastructures that provide services and resources to research communities assisting them in conducting research and making innovations. Research infrastructures are also crucial with re-

gard to collaboration and sustainability in the humanities and the cultural heritage sector. Moreover, they should help to overcome the digital divide by ensuring that also persons less versed in the use of digital technology can access and analyse material. Furthermore, research infrastructures should shed light on previously unnoticed data or topics. Additionally, findable and usable research infrastructures that follow good practice and standards are also crucial for public engagement. Related to research infrastructures are the digital commons that aim at making collections available and re-usable online free of charge, integrating various data sets, creating provenance and context for resources available online. Catalogues and databases should be interoperable, data consistent and data cleaning should be an inherent part of managing the digital commons (Hedges et al. 2019, 12-13).

2.4 Science Education

Science education has two major goals. The first goal is to enhance education so that citizens, including researchers and societal actors are equipped with the knowledge and skills to become RRI actors and participate in debates on and assume responsibility in the area of research and innovation. The second goal is to increase the interest in science among younger generations to either pursue a research career or to become a scientific citizen and contribute in a science-literate society (European Commission 2015, 6). This capacity building is necessary to foster change (EU 2012), connect science and society in order to pave the way for further innovation. This requires the interaction between the education and higher education system, funding of research and innovation, NGOs, civil society organisations, policy-makers, industry, professors, teachers, pupils and students as well as science centres or science museums to develop scientific citizenship and attract people to research and to develop RRI in university curricula (European Commission 2020e).

Science education is an inherent part of citizen humanities and, partly, also of the public humanities since the participants need some degree of factual or procedural knowledge when contributing to a project. In many citizen humanities projects or projects by cultural heritage institutions participants undergo training to be able to contribute to a project. This may be an introduction to tagging according to TEI principles, information about the history of a collection or metadata, transcription rules etc. Throughout a project, participants can further develop competences in certain fields (Dobreva, Azzopardi 2014, 451), including disciplinary, procedural or technological competences.

The citizen humanities can alter the relationship between members of the public, universities and cultural heritage institutions. Through

the citizen humanities, participants get an insight into academic research and the related processes. This does not only enhance academic literacy in general but also domain knowledge and transferable skills, such as critical and connected thinking, research and technological skills. The role of the humanities in imparting skills of critical thinking should not be underestimated in the digital age, which is shaped by information overload, fake news and post-factual tendencies. However, in the citizen humanities, learning is not unidirectional. Scholars can also learn from the participants which can improve their research and personal development (Heinisch et al. forthcoming) by being challenged in their traditional way of conducting research, being required to think out-of-the-box (Bonnefond, Riboli-Sasco, Sescousse 2015, 518). The DH, and online platforms in particular, allow a two-way dialogue, knowledge co-creation and community-building (Terras 2016, 421).

2.5 Public Engagement

Public engagement is at the interface between science, policy and society. It refers to the engagement and participation of all societal actors, including researchers, citizens, NGOs, civil society organisations, policy-makers and industry in research and innovation processes. This joint development of solutions should also help to tackle societal challenges based on representative concerns and common principles that are aligned to the needs, expectations and values of society (EU 2012). Therefore, multi-actor dialogues characterised by inclusion and participation are necessary. This means to embed public engagement in the research design and iteratively throughout the research process that ideally lead to co-created policy agendas and research and innovation outcomes. These outcomes should tackle societal challenges and be widely accepted. Here, CS is explicitly mentioned as a participatory research and innovation action. The benefits of public engagement according to the European Union are a scientifically literate society that can support democratic processes as well as research and innovation, injecting creativity and other perspectives in research and enhancing those outcomes of research and innovation that are relevant and desirable in society and that can tackle societal challenges. Furthermore, citizens should be engaged to contribute to policy or participatory foresight. Moreover, research and innovation policy can offer both knowledge and evidence that support thematic policies, for example on environment or health at different levels (European Commission 2020d). Suggested indicators of public engagement include policies, frameworks and regulations, events, initiatives and attention creation as well as competence building. Here, CS plays an important role as well (European Commission 2015, 6).

In its broadest sense, public engagement may refer to any type of interaction between science and society. Therefore, public engagement is not only at the core of the PH but also of the citizen humanities since they are nothing without their participants. The 'members of the public' or the 'citizens' in the citizen humanities may encompass different partners and groups, ranging from local communities, special interest communities, cultural heritage professionals, associations, elderly groups, trade unions, third-sector organisations, environmental teams, urban and rural councils, indigenous communities, engineers, government bodies, environmental impact consultants, public bodies, agencies, charities to school and university students, genealogists, NGOs and NPOs.

This includes approaches that are related to CS, but often not defined as such, for example, participatory health research, transdisciplinary research or public history. Moreover, public engagement as science-society interaction also encompasses science communication or science shops as well as open science (Pettibone, Vohland, Ziegler 2017, 12).

CS, also sometimes referred to as 'amateur science', has a strong focus on the inclusion of non-academic actors in academic research. In general, 'citizen' comprises anybody. However, people who are engaged in academic research as part of their profession or of professional training, for example, doctoral candidates, would rather not be regarded as citizen scientists (Pettibone, Vohland, Ziegler 2017, 12).

From the perspective of the public or engaged humanities, the crux in public engagement is to work at eye level and to foster partnerships that generate knowledge in a collaborative and reciprocal manner so that all participants, including researchers, students or communities are served (Jay 2012, 55).

Barriers to public engagement in public scholarship are that these types of research or collaboration are not recognised as activities helping to progress in the academic career. Furthermore, public engagement may also be just seen as outreach or service to the public (Jay 2012, 57). Moreover, collaborative research may not follow the model research process in the humanities, consisting of a linear research sequence, comprising the steps of finding a research question and a suitable methodology to answer this question, (collecting and) analysing data with the selected method, analysing the findings and disseminating them. In some cases, such as allotment projects, it may be hard to tell when the actual research starts and when it ends. Moreover, in collaborative projects, the research design, the methods and the outputs may be questioned, changed or adapted throughout the project. Additionally, the categories of 'expert', 'scholar' or 'activist' may be blurred in collaborative endeavours. Debates may develop in unexpected directions. Aspects of collaborative projects that are also related to ethics are the sustainability, potential impact

and the afterlife of a project. A project should be beneficial to all persons involved. Therefore, it may not end with a research article published in an academic journal by a researcher but may include other (non-proprietary, multi-authored) outputs, such as policy reports, transcription manuals, relationships or press releases, YouTube videos, blogs or social media discussions that are of different value for the persons involved. Another output can be the extrapolation of insights to non-academic contexts and audiences (The CRESC Encounters Collaborative 2013, 4-25).

A contested designation is 'citizen science' itself, since it may exclude all those persons that do not enjoy the official status of 'citizen'. Therefore, alternative terms that include all members of the public, such as 'community science' etc. have been proposed to avoid this type of exclusion. Nevertheless, as already addressed in the introduction, terms carry different connotations. Civic or community science is bottom-up science "initiated and driven by a group of participants who identify a problem that is a concern for them and address it using scientific methods and tools" (Haklay 2015, 15). Here, the community formulates the problem, collects and analyses data in collaboration with academics or research institutions.

The word 'science' in CS also raises the issue if participants in CS projects can be referred to as 'citizen scientists', since scientists have undergone professional training which enables them to address topics, apply methods and discuss theory. Therefore, there is a qualitative difference between the activities done by scholars and the activities scholars ask participants to perform. This is also the reason why some authors argue that participants in CS projects do not accomplish real scholarly work but rather perform auxiliary work that does often not go beyond data collection, data preparation or analysis (Terras 2016, 431). Here, public engagement can help promote research and extract free labour, but it also has the potential to empower participants and raise their motivation to further engage with a certain topic.

The notion of public engagement demonstrates a clear differentiation from outreach and service. On the one hand, outreach carries the connotation of a university that is privileged over the community and reinforces the view that universities are the only places of legitimate ownership and production of knowledge, where scholars are the guardians of expert knowledge (Gale, Carton 2005, 40). Service learning, on the other hand, addresses collaborations between the university and partners from the community to create intellectual projects that exert their effect in real life. Thus, students work on a project in the service of the community through which they should learn to apply research into practice (Jay 2012, 55).

Technology, that is at the core of the DH, offers many opportunities for humanities scholars and cultural heritage institutions. In ad-

dition to technologies used for collecting, analysing, storing and visualising (research) data, it also offers new ways of collaboration and communication, such as virtual reality, augmented reality, mobile apps or social media. However, creating usable technology in the citizen humanities that is appealing to different stakeholders that may have needs diverging from the scholars' needs is not always an easy task (Hedges et al. 2019, 11).

Technologies are an important means to solicit contributions from the public. This is exemplified by CS project directories or crowdsourcing platforms (Hedges et al. 2019, 11).

Thus, digital technology in general and the DH in particular, contribute to promote (volunteer) participation in CS. Centralised websites or digital platforms that list a wide variety of CS projects to which participants can contribute have become important means of participant recruitment (Colston, Vadjunec, Wakeford 2015, 67). Among these platforms are SciStarter, Zooniverse or national CS platforms, such as *Bürger schaffen Wissen* in Germany, *Österreich forscht* in Austria or *iedereenwetenschapper* in the Netherlands. Although members of the public can browse CS projects on these platforms and find a project that spark their interest, the projects usually have their own website or own app through which people can submit their contributions. For citizen humanities projects, this means that the project's website, the interfaces and the tools which the participants have to use should be characterised by a high usability. Here, the citizen humanities can draw on the experience gained from the DH in tool development on the one hand, and on the means of public engagement from the PH, on the other.

While digital technologies can also increase the digital divide, scholars principally see digital technologies as an opportunity that facilitates their research. Also, in citizen humanities projects, digital technologies play a central role since web-based technology allows a wide range of people to contribute to scholarly inquiries from the comfort of their homes. On the other hand, due to technological advances, such as artificial intelligence, some activities may become obsolete to which participants currently contribute, such as transcribing, tagging or pattern recognition (Heinisch et al. forthcoming).

2.6 Governance

Governance is the umbrella dimension that acts under the slogan "Design science for and with society" (EU 2012) and integrates all other RRI pillars. Governance covers processes, policies, rules and behaviour that affect the exertion of power. The EU has defined five requirements for good governance, including participation, accountability, openness, coherence and effectiveness. Institutional practic-

es and policy-making, i.e. governance in research and innovation, should become more accountable, transparent and inclusive. This refers to the policymakers' responsibility of preventing unethical or harmful developments. This should be achieved through RRI, which aligns innovation, science and society and fosters research and innovation that is more responsive to societal concerns, aspirations and needs. Therefore, topics under the governance pillar include incentives for responsible conduct both on an individual and an institutional level, the role of academic advice and expertise and the types of policies and processes needed to achieve RRI (European Commission 2020a). The overarching aim is to reach desirable and acceptable futures. This requires governance arrangements that are adaptable to the development of research and innovation, have to align with existing practices in research and innovation, share accountability and responsibility among actors and offer governance instruments for this shared responsibility (RRI Tools 2020). Since governance is the umbrella for all the other RRI pillars, it was already addressed before.

3 Case Study

On Everyone's Mind and Lips – German in Austria (abbreviated as IamDiÖ, `https://iam.dioe.at/`) is a citizen humanities project in the field of linguistics that was initiated by the Centre for Translation Studies at the University of Vienna, Austria, as an add-on to an already existing externally funded research project entitled *German in Austria. Variation – Contact – Perception* (`https://dioe.at/`). The latter addresses the variety and the change of the German language in Austria via the three aspects of variation, contact and perception, thereby investigating the ways the German language is used and perceived in Austria and showing the influence of other languages on the German language in Austria. Additional funding was acquired for a citizen science project (iam.dioe.at) that follows different approaches to citizen science. The first approach, the 'Question of the Month', consists of co-creation, i.e. the participants select a research question and, with the help of the researchers, decide on a method, collect and analyse the data and publish the results. The second approach is called 'Linguistic Treasure Hunt' addressing the Austrian linguistic landscape. Participants take pictures of written text (in any language or language variety) in the public sphere, such as on the streets or in public buildings and annotate them by specifying the geographical location, the language(s) and language variety, the medium, the context, function etc. While the Question of the Month concentrates on involving members of the public in the entire research process, the Linguistic Treasure Hunts rather focus on data collection and an initial data analysis by the participants. In the following section Iam-

DiÖ is analysed according to the RRI principles introduced above.

In the field of ethics, since the project participants are using language, which is at the core of linguistics, citizen linguistics raises the issue of where to draw the line between research subject and co-researcher. However, since IamDiÖ participants performed scholarly tasks, such as data collection and analysis, themselves, they were not the subjects of investigation but co-researchers.

Related to legal issues, the General Data Protection Regulation and intellectual property rights were affecting IamDiÖ. Participants could submit their Questions of the Month via a form on the project website that requires participants to enter personal data. Intellectual property rights were considered for the blog entry in which they answer their research questions. For the Linguistic Treasure Hunts, the participants needed to register via an app to upload and tag their photos.

Since the RRI principle of ethics also refers to the societal relevance and ethical acceptability of research, participants in several linguistic treasure hunts were surveyed afterwards. Although ethical acceptability was not part of the questionnaire, the respondents indicated an increased awareness for linguistic landscapes and expressed their motivation to continue their participation. In an informal meeting, one person also indicated that he integrated the search for written text in the public sphere in his daily routine and used breaks and daily routes to contribute to research (i.e. a greater good).

Although the project was aimed at dialogue at eye level, giving all persons an equal voice, seeing participants as co-researchers and using informal ways of communication, including being on a first-name basis with the participants (which is rather unusual in the communication with unfamiliar persons in German), IamDiÖ could not (entirely) overcome the hierarchies and the deficit model. On the one hand, this may be due to the project design of the Linguistic Treasure Hunts, where the researchers define the design, specify the categories of analysis according to their interests and provide participants with instructions. Here, participants hardly have a say. However, a similar observation was made for the Question of the Month. Only a small number of participants could be reached via social media or the website. Interestingly, the collaboration worked best in situations with face-to-face communication compared to communication through online means. The majority was reached through face-to-face interactions during science communication festivals, where visitors were eagerly asking questions about the topic of German in Austria but were not willing to find an answer to their question on their own. The reasons mentioned by some visitors suggest that the deficit model still prevails, i.e. "You are the researchers. You should know the answer/ You should find the answer", clearly specifying who is knowledge producer or knowledge receiver. Another reason mentioned for not partic-

ipating was a lack of time to delve into a research question. Although persons may be personally affected by their question, e.g. "Do dialects disappear?" if they are speakers of a dialect in Austria and may consider the question to be of societal and personal relevance, this did not result in (further) participation in the project. This, again, may be due to the setting of the personal encounter during a science festival where researchers present themselves as experts and where they could not provide prospective participants with the necessary information on the project and could not emphasise the potential benefits, such as personal development, receiving training or being part of a community, which would be required to build trust.

Another ethical concern is the sustainability of the project since it is only funded for a period of two years. This is a rather short period to build a community, establish partnerships and trust. Although a small community could be built and their contributions feed into an openly accessibly research infrastructure that makes their contributions re-usable, they may be frustrated if, after two years, their contributions and their interests are no longer needed, not used or not acknowledged.

Other ethical issues raised during the project were the use of incentives, such as prizes for Linguistic Treasure Hunts, and the degree of voluntariness as well as the notion of 'citizens', if university students receive bonus points for a course if they participate in research fields in which they are actually trained in.

Regarding the RRI pillar of gender equality, IamDiÖ did not achieve a gender balance, neither among the core project team nor among the participants, both dominated by females. Therefore, the project may not consider the needs and attitudes of all genders. Moreover, the gender dimension is only addressed indirectly in research, e.g. if participants raise these issues as part of a Question of the Month.

With regard to the open access RRI pillar, the majority of the material developed as part of IamDiÖ is made openly accessible, either via the website, such as educational and training material, answered Questions of the Month as blog entries or photos from Linguistic Treasure Hunts on the relevant project website and in the app. Moreover, the results also feed into a research infrastructure addressing the topic of German in Austria. Academic publications, such as journal articles, are also published in an open-access format. However, the adherence to the FAIR principles and the visibility of the participants' contributions and ideas could be (further) increased. IamDiÖ strongly relied on digital (in many cases also visual) means of collaboration and communication that may exclude certain groups of people, e.g. people with disabilities or people not having internet access or not being versed in using digital technologies.

With respect to the science education RRI pillar, IamDiÖ enhanced education among citizens. The humanities, and especially the human-

ities conducting fundamental research, do not receive a lot of recognition, which held partially also true for IamDiÖ, when participants questioned the need for research in the project's topics. While the volunteers were strongly participating in debates in the area of research, they strongly relied on language myths and their personal experience and opinions. Therefore, the project was able to present the academic foundation and results (underlying or contradicting their experience or opinions). This showed that equipping societal actors with the relevant knowledge and skills is crucial for making informed decisions. Although the Question of the Month aimed at both, increasing the participants' (factual and procedural) knowledge and academic skills to contribute to a science-literate society, the initial concept of the Question of the Month did not meet the expectations of the public. Whether the Question of the Month could increase academic literacy and alter the relationship between the university and the public requires further research. As part of the science education element, scholars and participants learned from each other, since the researchers saw the societal relevance of their research topics exemplified by the number of questions raised by members of the public and they also saw research gaps when volunteers raised issues that are not or under-explored in academic research. Although the digital humanities offer various means of communication, for IamDiÖ personal dialogue resulted in the most fruitful discussions.

Regarding the RRI pillar of public engagement, the research design for the Question of the Month changed drastically during the process based on the interaction with participants. The initial idea that participants do not only raise but also answer their questions could not be implemented because the number of persons who were willing to go through the entire research process was too small. Therefore, the Question of the Month was changed according to the feedback from the visitors at science festivals. Although persons could still answer their own research questions, IamDiÖ researchers also answered some questions from the participants if there were already academic findings available. Thus, the Question of the Month, which was initially intended as co-creation rather became a science communication exercise. Adaptations of the categories of analysis and their explanations were also made for the linguistic treasure hunt based on the participants' feedback. However, these were only minor incremental changes. Moreover, in linguistics, the boundaries between expert and non-expert are blurred when it comes to the use of a (local) language (variety), when participants can often draw on considerable practical experience. Additionally, by selecting blog posts and social media for disseminating and voting on the Question of the Month, IamDiÖ aimed at reaching a broader audience. Although IamDiÖ was listed on the Austrian citizen science platform, where CS projects that meet certain quality criteria present them-

selves to attract participants, this was not the most successful recruitment strategy. This runs counter to the argument that digital platforms help promote volunteer participation in CS.

Since almost no background information on the participants was collected in the project, the 'citizens' cannot be specified in more detail. However, IamDiÖ will directly approach special interest associations, e.g. dialect associations, in the future for conceptualising the research design, frameworks, events, competence building and policy actions with them.

Regarding the RRI pillar of governance, IamDiÖ was designed as an open project allowing anybody to participate in linguistic research. However, since the Question of the Month could not be implemented as planned, also the project governance was affected. Although IamDiÖ strived for transparency and being responsive to societal needs and concerns, inclusion and shared responsibility could not be achieved.

To sum up, although IamDiÖ aimed at co-creation (for the Question of the Month) and aligns well with several RRI pillars, the initial plan could not be implemented for reasons that still need further investigation, but that may be related to the unconventional research project design, where many aspects are unclear in the beginning, such as the research question, the methods and the outcomes. This fuzziness is difficult to communicate and to relate to the participants' life worlds, i.e. making the project useful (enough) and meaningful to its participants. Moreover, it seems that the deficit model prevails in the participants' minds impeding them from taking agency and responsibility in the project. Interestingly, the digital means of collaboration were less successful than face-to-face encounters. This case study also demonstrates that "sustainable co-creation can only emerge after considerable time and effort has been made to cultivate a spirit of trust and reciprocity grounded in an embodied respect for cultural knowledge and experience" (Miller, Little, High 2017, 10).

4 Discussion

This study set out with the aim of assessing the fusion of the digital and public humanities in the form of the citizen humanities. Although the definition of citizen science is a contested one, it shows clear links to RRI.

Although the citizen humanities can be considered a fusion between the DH and PH, they go one step further. Nevertheless, aspects that all three forms of the humanities (DH, PH and citizen humanities) have in common are their object of research, namely human culture, as well as their roots in scholarship, their potential for the advancement of research and contribution to knowledge pro-

duction. Moreover, these forms of the humanities continue the tradition of "critical thinking, interdisciplinarity, debate over values, and the posing of profound philosophical questions typical of humanities scholarship" (Jay 2012, 53). While the 'traditional' humanities significantly differ from the tools and methods used in the research of the three forms of the humanities presented in this paper, the DH, PH and citizen humanities have certain tools and methods in common, since they all heavily rely on the use of digital technology. However, going digital is not the only way to conduct citizen humanities, since volunteers who travel to archives, collect documents, gather local and special knowledge or who participate in field surveys do not necessarily require digital infrastructures and tools provided by the DH (Heinisch et al. forthcoming).

A major difference between these three forms of the humanities is public engagement. While the degree of public engagement might differ significantly in PH or citizen humanities projects, the DH are, in general, not aimed at non-academics (beyond outreach and dissemination). However, the DH can be regarded as auxiliary (discipline) for the PH and the citizen humanities providing tools and methods that support them, or that make them possible, in the first place. Digital technologies, including the Internet, lead to spaces that are inhabited by both non-academic and academic communities and which provide the rooms for connecting academic research to public communities (Dunn, Hedges 2012, 3).

The *Digital Humanities Manifesto 2.0* creates connections to the citizen humanities since it acknowledges that digital tools, media and techniques changed the generation and dissemination of knowledge, including new ways of scholarly discourse in which universities are no longer the sole stewards of culture or knowledge. Although the DH seem to have a focus on quantitative rather than qualitative work, also the experiential, interpretative, generative and emotive nature of the DH should be taken into account (DH 2020). While the quantitative aspect of the DH would rather foster crowdsourcing in the humanities and cultural heritage institutions, the qualitative character would help boost co-created approaches in the citizen humanities as well. Moreover, since universities are no longer the sole producers of knowledge, also other types of knowledge, e.g. local knowledge could gain a foothold as well.

Moreover, the *Manifesto* stresses that interdisciplinarity or transdisciplinarity would require changes in the DH themselves, including methods, practice and output. In addition, it fosters the democratisation of culture and scholarship (DH 2020). This again shows that the DH build bridges not only to other disciplines but also beyond academia, which includes public engagement in humanistic inquiry, and thus the citizen humanities. A direct link to CS can also be found in the request to democratise scholarship which was also at the heart

of Irwin's (1995) understanding of citizen science. While the *Manifesto* highlights the multi-purposing of humanistic knowledge and the creation of bigger pictures and co-creation (DH 2020), it still emphasises the role of experts. Co-creation rather means teamwork among scholars although the Manifesto also sees the DH "as an umbrella under which to group both people and projects seeking to reshape and reinvigorate contemporary arts and humanities practices, and expand their boundaries" (DH 2020).

The contributions of the DH to the citizen humanities are manifold. While the humanities primarily produce textual outcomes, the DH allow for additional media, such as platforms, images and different types of visualisations. Since the DH rely on computational methods, larger amounts of data can be collected, stored, analysed, shared and presented, also automatically or semi-automatically. This requires or opens up new ways of collaboration and publication, not only among academics but also with non-academics, as it is the case in the citizen humanities (Heinisch et al. forthcoming).

The PH contribute to the citizen humanities in different ways: first, citizen humanities as well as PH require the establishment of partnerships, dialogue and trust between academic researchers and volunteers or participants. Transparency and feedback are key to citizen humanities projects. This may also require to make volunteers advocates, not only for the materials they are working on and the project but also for the discipline and the humanities as a whole (Terras 2016, 431). Citizen humanities mean a shift in the scholars' thinking – away from what citizens can do for science to how to bring together humanities disciplines, DH (and public humanities) to conduct citizen humanities (Belknap 2015).

Kimura and Kinchy (2016, 339) found seven distinctive virtues of CS, encompassing the increase in data available for research, the enhancement of public understanding of research as well as of community capacity to address environmental issues, the formation of more equal university-community relationships, the closing of knowledge gaps and putting official accounts under scrutiny, driving policy change and catching polluters (the latter rather specific to natural science projects).

This demonstrates that the aim of 'going public' may not be to reach a high number of people but rather to influence only a small number of stakeholders or policy makers. Therefore, it is necessary to tailor communication to the target audience, including the language, the form and the aesthetics (Miller, Little, High 2017, 8).

Despite the benefits and the promises offered by the citizen humanities, not all scholarly projects are suited for public engagement and not all processes and decisions within a project qualify to be opened to the public. This may be due to the complexity, the sensitivity or the specialisation of the object of investigation or the research

project (Davidson 2009). Therefore, further research should be undertaken to investigate the different relations and interrelations, boundaries and overlaps between other forms of the humanities.

The citizen humanities (as well as DH and PH) are influenced by framework conditions. Corporate practices, defunding in universities, power disparities as well as a general lack of status and (competition for) resources characterise humanities disciplines (at universities). Especially, collaboration with members of the public is often interdisciplinary work for which it may be more difficult to receive funding. Therefore, institutional precarity, institutionalisation, disciplinarity and formalisation can be barriers and informal labour may be needed for transformation (Desai, Murphy 2018, 26-39). Moreover, funding bodies may consider citizen science as a means to bridge gaps in research funding, since the labour of the volunteers participating in CS projects is 'for free'. However, this perspective is neglecting the amount of work needed to run a CS project, including participant recruitment, communication, preparation of training materials, events, protocols, software development etc. Since the citizen humanities may not always address an urgent societal need or concern (compared to topics such as air pollution or ethnic segregation that affect persons directly), it may be difficult to argue for an immediate impact or effect.

Additionally, actors may not always participate in collaborative research freely, but out of the need to respond to problems (such as ethnic segregation and racism), budget cuts or limited capacities (The CRESC Encounters Collaborative 2013, 25).

Furthermore, the use of digital technology in citizen humanities may increase the digital divide. As illustrated by the case study, digital technology worked best in combination with more traditional forms of interaction. Although the citizen humanities and cultural heritage institutions make use of latest technologies to remain relevant, reach a broader audience and make history more meaningful, the digital divide also needs to be considered in university-community partnerships (Hurley 2016, 70).

Tensions in the citizen humanities are similar to those found in other collaborative endeavours, including institutional imperatives, (power) inequalities, ownership (of the project and the results), diverging interests and different agendas of the persons involved in a project. For example, researchers may wish to publish an academic paper (keyword: 'publish or perish') while city councils may require a policy paper and local communities may wish for tangible change in their environment which may run counter to the other actors' objectives. Collaborative projects are also subject to the tension of power relations: who participates under which conditions and whose voices are prioritised. If there are hardened fronts, researchers may also (be expected to) act as unbiased and objective mediators. Col-

laboration may also be characterised by opportunism as well as the ambition of balancing ethics and politics. To be a success, participants need a common interest, a common problem and a common politics (The CRESC Encounters Collaborative 2013, 4-26). Researchers are, in this respect, also cultural producers and advocates for social change (Miller et al. 2017, 14). Therefore, impact, benefits and relationships may take various forms and may rely on reciprocal 'gifts'. These 'gifts' can be reports or the acquisition of expertise through training. Therefore, the mutual benefits of collaboration can also be expertise or information. However, different actors may have different benefits, ranging from monetary value, social networks to prestige or legitimation (The CRESC Encounters Collaborative 2013, 17-24). Mutually meaningful and mutually beneficial outcomes can be related to RRI itself: achieving acceptable, desirable and sustainable research outcomes and processes that align with the values, expectations and needs of society, and that help to address societal challenges.

Therefore, citizen science, including the citizen humanities, may cause cultural change (both in academia and in society) and lead to new ways of thinking. As illustrated by the case study, participants have to understand that they are no longer the subject of investigation but conduct research (partially) themselves. Depending on the governance aspect of a citizen humanities project, participants may also help to design the project itself and assume responsibility for the project and its outcomes.

What differentiates the citizen humanities from other types of public engagement then? The citizen humanities have a focus on knowledge (co-)production (and the advancement of research). Although grassroots activities (where the community identifies a problem and the knowledge of the local community, in addition to scientific findings, are used to solve the problem) with a research objective also fall under the citizen humanities, the citizen humanities are often driven by professional (university) researchers and cultural heritage institutions.

While equity, inclusiveness, social justice and well-being may be the goal (or one of the goals) of citizen humanities projects, they are not a prerequisite. This also means that the citizen humanities (and the produced knowledge) are not necessarily transformative or empowering. To exert impact, knowledge creation needs to be context-relevant (Leadbeater, Banister, Marshall 2011, 9-10).

5 Conclusion

The citizen humanities engage members of the public in humanistic research with the aim of (co-)producing knowledge, or even change.

The citizen humanities are related to responsible research and innovation. Although the digital and public humanities helped to pave the way for the citizen humanities, the citizen humanities are more than a fusion between them. These findings raise intriguing questions regarding the nature of the citizen humanities and their relation to other forms of public engagement and academic research. However, the line between the different types of the humanities, including academic humanities, public humanities, digital humanities, public digital humanities and citizen humanities can get fuzzy.

Bibliography

Bela, G.; Peltola, T.; Young, J.C.; Balázs, B.; Arpin, I.; Pataki, G.; Hauck, J.; Kelemen, E.; Kopperoinen, L.; Herzele, A.; Keune, H.; Hecker, S.; Suškevičs, M.; Roy, H.E.; Itkonen, P.; Külvik, M.; László, M.; Basnou, C.; Pino, J.; Bonn, A. (2016). "Learning and the Transformative Potential of Citizen Science". *Conservation Biology*, 30(5), 990-9.

Belknap, G. (2015). "From Citizen Science to Citizen Humanities – 19th Century History in the Digital Age". https://bit.ly/34ifnfl.

Bonnefond, M.; Riboli-Sasco, L.; Sescousse, G. (2015). "Repainting citizen science". *Science (New York)*, 350(6260), 518. https://doi.org/10.1126/science.350.6260.518-b.

Bonney, R.; Ballard, H.; Jordan, R.; McCallie, E.; Phillips, T.; Shirk, J.; Wilderman, C.C. (2009). "Public Participation in Scientific Research: Defining the Field and Assessing Its Potential for Informal Science Education. A CAISE Inquiry Group Report". Center for Advancement of Informal Science Education. http://files.eric.ed.gov/fulltext/ED519688.pdf.

Boyte, H.C.; Farr, J. (1997). The Work of Citizenship and the Problem of Service-Learning. Civic Engagement 2. https://digitalcommons.unomaha.edu/slceciviceng/2.

Brennan, S. (2016). "Public, First". Gold, M.K.; Klein, L.F. (eds), *Debates in the Digital Humanities*. Minneapolis: University of Minnesota Press, 384-90.

Burdick, A.; Drucker, J.; Lunenfeld, P.; Presner, T.; Schnapp, J. (2012). *Digital_humanities*. Edited by A. Burdick, J. Drucker, P. Lunenfeld, T. Presner, J. Schnapp. Cambridge (MA): Cambridge: MIT Press.

Carton, E. (2009). "Keyword: Public Humanities. Interview by Kevin Bott". Imagining America, 12, Spring, 11-12.

Cole, C.C. (1994). "Public History: What Difference Has It Made?". *The Public Historian*, 16(4), 9-35. https://doi.org/10.2307/3378008.

Colston, N. M.; Vadjunec, J. M.; Wakeford, T. (2015). "Exploring the Entry Points for Citizen Science in Urban Sustainability Initiatives". *Current Opinion in Environmental Sustainability*, 17, 66-71. https://doi.org/10.1016/j.cosust.2015.11.006.

Cox, J.; Tilton, L. (2019). "The Digital Public Humanities: Giving New Arguments and New Ways to Argue". *Review of Communication*, 19(2), 127-46. https://doi.org/10.1080/15358593.2019.1598569.

Davidson, C. (2009). "Calling Citizen Humanists". HASTAC. https://www.hastac.org/blogs/cathy-davidson/2009/05/27/calling-citizen-humanists.

Desai, J.; Murphy, K.P. (2018). "Subjunctively Inhabiting the University". *Critical Ethnic Studies*, 4(1), 21. https://doi.org/10.5749/jcritethnstud.4.1.0021.

DH (2020). "The Digital Humanities Manifesto 2.0". http://manifesto.humanities.ucla.edu/2009/05/29/the-digital-humanities-manifesto-20/.

Dobreva, M.; Azzopardi, D. (2014). "Citizen Science in the Humanities: A Promise for Creativity". Papadopoulos, G.A. (eds), *Proceedings of the 9th International Conference on Knowledge, Information and Creativity Support Systems* (Limassol, Cyprus, November 6-8, 2014). Cyprus: Cyprus Library, 446-51.

Dunn, S.; Hedges, M. (2012). "Crowd-Sourcing Scoping Study. Engaging the Crowd with Humanities Research". Arts & Humanities Research Council. https://stuartdunn.files.wordpress.com/2013/04/crowdsourcing-connected-communities.pdf.

ECSA (2020). "ECSA's Characteristics of Citizen Science". https://doi.org/10.5281/zenodo.3758668.

Ellison, J. (2013). "Guest Column: The New Public Humanists". *PMLA*, 289-98.

enRRICH (2016). EnRRICH – Enhancing Responsible Research and Innovation through Curricula in Higher Education. https://www.das-wissen-teilen.de/en/unsere-projekte/enrrich/.

EU (2012). Responsible Research and Innovation. Europe's Ability to Respond to Societal Challenges. https://ec.europa.eu/research/swafs/pdf/pub_public_engagement/responsible-research-and-innovation-leaflet_en.pdf.

European Citizen Science Association (2015). "Ten Principles of Citizen Science". http://ecsa.citizen-science.net/sites/default/files/ecsa_ten_principles_of_citizen_science.pdf

European Commission (2015). Indicators for promoting and monitoring Responsible Research and Innovation. Report from the Expert Group on Policy Indicators for Responsible Research and Innovation. https://ec.europa.eu/research/swafs/pdf/pub_rri/rri_indicators_final_version.pdf#view=fit&pagemode=none.

European Commission (2020a). "Governance". https://ec.europa.eu/research/swafs/index.cfm?pg=policy&lib=governance.

European Commission (2020b). "Open Science (Open Access)". https://ec.europa.eu/programmes/horizon2020/node/1031.

European Commission (2020c). "Promoting Gender Equality in Research and Innovation". https://ec.europa.eu/programmes/horizon2020/node/797.

European Commission (2020d). "Public Engagement in Responsible Research and Innovation". https://ec.europa.eu/programmes/horizon2020/node/766.

European Commission (2020e). "Science Education". https://ec.europa.eu/programmes/horizon2020/node/795.

Fitzgerald, H.E.; Primavera, J. (eds) (2013). "Going public. Civic and Community Engagement". East Lansing: Michigan State University Press.

Gale, S.; Carton, E. (2005). "Toward the Practice of the Humanities". *The Good Society*, 14(3), 38-44. https://doi.org/10.1353/gso.2006.0006.

Haklay, M. (2013). "Citizen Science and Volunteered Geographic Information – Overview and Typology of Participation". Sui, Elwood, Goodchild 2013, https://doi.org/10.1007/978-94-007-4587-2_7.

Haklay, M. (2015). *Citizen Science and Policy: A European Perspective*. Woodrow Wilson International Center for Scholars.

Hedges, M.; Dunn, S. (2018). *Academic Crowdsourcing in the Humanities. Crowds, Communities and Co-production*. Cambridge (MA): Elsevier.

Hedges, M.; Stuart, D.; Tzedopoulos, G.; Bassett, S.; Garnett, V.; Giacomi, R.; Sanesi (2019). "Digital Humanities Foresight: the Future Impact of Digital Methods, Technologies and Infrastructures". DARIAH-DE Working Papers Nr. 40. http://nbn-resolving.de/urn:nbn:de:gbv:7-dariah-2019-12-3.

Heigl, F.; Dörler, D.; Bartar, P.; Brodschneider, R.; Cieslinski, M.; Ernst, M.; Fritz, S.; Greilhuber, I.; Hatlauf, J.; Hecker, S.; Hübner, T.; Kieslinger, B.; Kraker, P.; Krennert, T.; Oberraufner, G.; Paul, K.T.; Tiefenthaler, B.; Vignoli, M.; Walter, T.; Würflinger, R.; Zacharias, M.; Ziegler, D. (2018). *Quality Criteria for Citizen Science Projects on Österreich forscht | Version 1.1*. https://doi.org/10.31219/osf.io/48j27.

Heigl, F.; Kieslinger, B.; Paul, K.T. Uhlik, J.; Dörler, D. (2019). "Opinion: Toward an International Definition of Citizen Science". *Proceedings of the National Academy of Sciences of the United States of America*, 116(17), 8089-92. https://doi.org/10.1073/pnas.1903393116.

Heinisch, B.; Oswald, K.; Weißpflug, M.; Shuttleworth, S.; Belknap, G. (forthcoming). "Citizen Humanities". Vohland, K.; Land, A.; Ceccaroni, L. et al. (eds), *The Science of Citizen Science*. Chams: Springer.

Henke, J.; Pasternack, P.; Schmid, S. (2018). *Mission, die dritte. Die Vielfalt jenseits hochschulischer Forschung und Lehre: Konzept und Kommunikation der Third Mission*. Berlin, BWV Berliner Wissenschafts-Verlag.

Hughes, L.; Constantopoulos, P.; Dallas, C. (2016). "Digital Methods in the Humanities". Schreibman, Siemens, Unsworth 2016, 150-70.

Hurley, A. (2016). "Chasing the Frontiers of Digital Technology". *The Public Historian*, 38(1), 69-88. https://doi.org/10.1525/tph.2016.38.1.69.

Irwin, A. (1995). *Citizen Science. A Study of People, Expertise and Sustainable Development*. London: Routledge.

Jay, G. (2012). "The Engaged Humanities: Principles and Practices for Public Scholarship and Teaching". *Journal of Community Engagement and Scholarship*, 3(1), 51-63.

Kenny, S.; MacGrath, B. (eds) (2018). *The Routledge Handbook of Community Development. Perspectives from Around the Globe*. New York, NY: Routledge.

Kieslinger, B.; Schäfer, T.; Heigl, F.; Dörler, D.; Richter, A.; Bonn, A. (2018). "Evaluating Citizen Science. Towards an Open Framework". Hecker, S.; Haklay, M.; Bowser, A. et al. (eds). *Citizen Science. Innovation in Open Science, Society and Policy*. London: UCL Press, 81-95.

Kimura, A.H.; Kinchy, A. (2016). "Citizen Science. Probing the Virtues and Contexts of Participatory Research". *Engaging Science, Technology, and Society*, 2, 331. https://doi.org/10.17351/ests2016.99.

Kullenberg, C.; Kasperowski, D. (2016). "What Is Citizen Science?--A Scientometric Meta-Analysis". *PloS one*, 11(1), 1-16. https://doi.org/10.1371/journal.pone.0147152.

Leadbeater, B.J.; Banister, E.M.; Marshall, E.A. (2011). "How-What-We-Know-Becomes-More-Widely-Known Is Context Dependent and Culturally Sensitive". Banister, E.M.; Leadbeater, B.; Marshall, A. (eds), *Knowledge Translation in Context. Indigenous, Policy, and Community Settings*. Toronto: University of Toronto Press, 1-12.

Lindner, R.; Goos, K.; Güth, S.; Som, O.; Schröder, T. (2016). "'Responsible Research and Innovation' als Ansatz für die Forschungs-, Technologie- und Innovationspolitik – Hintergründe und Entwicklungen". https://www.tab-beim-bundestag.de/de/pdf/publikationen/berichte/TAB-Hintergrundpapier-hp022.pdf.

Miller, E.; Little, E.; High, S.C. (2017). *Going Public. The Art of Participatory Practice*. Toronto: University of British Columbia Press.

Miller-Rushing, A.J.; Primack, R.B.; Bonney, R.; Albee, E. (2020). "The History of Citizen Science in Ecology and Conservation". Lepczyk, C.A.; Boyle, O.D.; Vargo, T.L.V. et al. (eds), *Handbook of Citizen Science in Ecology and Conservation*. Oakland: California, University of California Press, 17-24.

Miller-Rushing, A.; Primack, R.; Bonney, R. (2012). "The History of Public Participation in Ecological Research". *Frontiers in Ecology and the Environment*, 10(6), 285-90. https://doi.org/10.1890/110278.

Molas-Gallart, J.; Salter, A.; Patel, P.; Scott, A.; Duran, X. (2002). *Measuring Third Stream Activities. Final Report to the Russell Group of Universities*. Science and Technology Policy Research.

Pettibone, L.; Vohland, K.; Bonn, A.; Richter, A.; Bauhus, W.; Behrisch, B.; Borcherding, R.; Brandt, M.; Bry, F.; Dörler, D.; Elbertse, I.; Glöckler, F.; Göbel, C.; Hecker, S.; Heigl, F.; Herdick, M.; Kiefer, S.; Kluttig, T.; Kühn, E.; Kühn, K.; Oswald, K.; Röller, O.; Schefels, C.; Schierenberg, A.; Scholz, W.; Schumann, A.; Sieber, A.; Smolarski, R.; Tochtermann, K.; Wende, W.; Ziegler, D. (2016). Citizen Science für alle. Eine Handreichung für Citizen Science-Beteiligte. Bürger Schaffen Wissen (GEWISS)-Publikation. Deutsches Zentrum für Integrative Biodiversitätsforschung (iDiv) Halle-Jena-Leipzig, Helmholtz-Zentrum für Umweltforschung – UFZ, Leipzig; Berlin-Brandenburgisches Institut für Biodiversitätsforschung (BBIB), Museum für Naturkunde (MfN) – Leibniz-Institut für Evolutions- und Biodiversitätsforschung. http://buergerschaffenwissen.de/sites/default/files/assets/dokumente/gewiss_citscifueralle_handreichung_web_0.pdf.

Pettibone, L.; Vohland, K.; Ziegler, D. (2017). "Understanding the (Inter)disciplinary and Institutional Diversity of Citizen Science. A Survey of Current Practice in Germany and Austria". *PloS one*, 12(6), e0178778. https://doi.org/10.1371/journal.pone.0178778.

Reiheld, A.; Gay, P.L.(2019). "Coercion, Consent, and Participation in Citizen Science". https://arxiv.org/abs/1907.13061.

Riesch, H.; Potter, C. (2014). "Citizen Science as Seen by Scientists: Methodological, Epistemological and Ethical Dimensions". *Public Understanding of Science*, 23(1), 107-20. https://doi.org/10.1177/0963662513497324.

Rome Declaration on Responsible Research and Innovation in Europe (2014). https://ec.europa.eu/research/swafs/pdf/rome_declaration_RRI_final_21_November.pdf.

RRI Tools (2020). "What is RRI?". https://www.rri-tools.eu/about-rri.

RRI-Plattform Österreich (2020). Policy Brief: Responsible Research and Innovation in Österreich. https://www.rri-plattform.at/wp-content/uploads/2020/03/RRI-Policy-Brief-RRI-Plattform.pdf.

Sample, M. (2013). "The Digital Humanities is not about Building, it's about Sharing". Terras, Nyhan, Vanhoutte 2013, 255-7.

Schomberg, R. von (2012). *Prospects for Technology Assessment in a Framework of Responsible Research and Innovation*. VS Verlag für Sozialwissenschaften.

Schreibman, S.; Siemens, R.G.; Unsworth, J. (eds) (2016). *A New Companion to Digital Humanities*. Chichester: Wiley Blackwell,

Serrano Sanz, F.; Holocher-Ertl, T.; Kieslinger, B.; García Sanz, F. ; Silva, C.G. (2014). "White Paper on Citizen Science for Europe". http://www.socientize.eu/sites/default/files/white-paper_0.pdf

Shirk, J.L.; Ballard, H.L.; Wilderman, C.C.; Phillips, T.; Wiggins, A.; Jordan, R.; McCallie, E.; Minarchek, M.; Lewenstein, B.V.; Krasny, M.E.; Bonney, R. (2012). "Public Participation in Scientific Research. A Framework for Deliberate Design". *Ecology and Society*, 17(2). https://doi.org/10.5751/ES-04705-170229.

Siemens, L. (2010). "Time, Place and Cyberspace: Foundations for Successful e-Research Collaboration". Anandarajan, M.; Anandarajan, A. (eds), *E-Research Collaboration. Theory, Techniques and Challenges*. Berlin: Springer, 35-48.

Silvertown, J. (2009). "A New Dawn for Citizen Science". *Trends in Ecology & Evolution*, 24(9), 467-71. https://doi.org/10.1016/j.tree.2009.03.017.

Sui D.; Elwood S.; Goodchild M. (eds) (2013). *Crowdsourcing Geographic Knowledge*. Dordrecht: Springer. https://doi.org/10.1007/978-94-007-4587-2.

Tanner, S. (2015). "Crowdsourcing is Dead - Long Live Citizen Humanities". http://simon-tanner.blogspot.co.at/2015/09/crowdsourcing-is-dead-long-live-citizen.html.

Terras, M.; Nyhan, J.; Vanhoutte, E. (eds) (2013). *Defining Digital Humanities. A Reader.* Farnham: Ashgate Publishing Ltd,

Terras, M. (2016). "Crowdsourcing in the Digital Humanities". Schreibman, Siemens, Unsworth 2016, 420-38.

The CRESC Encounters Collaborative (2013). "(Un)Doing Collaboration: Reflections on the Practices of Collaborative Research". http://research.gold.ac.uk/8850/1/CRESCEncountersCollective(Un)doingCollaborationWP127_0.pdf.

Tweddle, J.C.; Robinson, L.D.; Pocock, M.J.O.; Roy, H.E. (2012). "Guide to Citizen Science: Developing, Implementing and Evaluating Citizen Science to Study Biodiversity and the Environment in the UK". http://www.ceh.ac.uk/sites/default/files/citizensscienceguide.pdf.

Unger, H. von (2014). *Partizipative Forschung. Einführung in die Forschungspraxis*. Wiesbaden: Springer.

Vanhoutte, E. (2013). "The Gates of Hell: History and Definition of Digital | Humanities | Computing". Terras, Nyhan, Vanhoutte 2013, 119-56.

Wickman, M. (2016). "What Are the Public Humanities? No, Really, What Are They?". *University of Toronto Quarterly*, 85(4), 6-11. https://doi.org/10.3138/utq.85.4.6.

Wickman, M.; Browne, J. (2014). "CHCI Public Humanities Survey–Draft Report. BYU Humanities Center". https://humanitiescenter.byu.edu/files/2016/06/CHCI-Public-Humanities-Survey-Report.pdf.

e-ISSN 2724-3923

magazén
Vol. 1 – Num. 2 – December 2020

A Brief History of the Theory and Practice of Computational Literary Criticism (1963-2020)

Chris Beausang
Maynooth University, Ireland

Abstract This paper will construct a history of computational literary criticism (CLS) which has engaged statistical methods by providing an historical account of the journal articles as well as other publications which have advanced the field to the most significant extent since 1963. This paper divides the history of CLS into three distinct epochs, within each of which the methods and theories CLS scholars utilise undergo significant qualitative transformation. The decisive factor in each of these epochs is CLS' relationship to traditional literary criticism. Partly as a result of this, CLS scholarship initially cleaves to organic theories of literary style and adopts a highly polemicised opposition to then-regnant post-structuralist theories of authorship.

Keywords Computational literary criticism. Literary criticism. Digital humanities. Delta. Machine learning.

Summary 1 Introduction. – 2 Embryonic CLS (1963-1979). – 3 PCA & Proto-Delta (1980-1990). – 4 Delta, Results and Prospects (2000-2020). – 5 Conclusion.

Edizioni Ca'Foscari

Peer review

Submitted	2020-10-09
Accepted	2020-11-02
Published	2020-12-22

Open access

Citation Beausang, C. (2020). "A Brief History of the Theory and Practice of Computational Literary Criticism (1963-2020)". *magazén*, 1(2), 181-202.

DOI 10.30687/mag/2724-3923/2020/02/002

1 Introduction

The heterogeneity of computational literary studies (CLS) can render it difficult to map. Its being integral to the digital humanities has the consequence that it would not be uncommon for an article or a chapter which marks a significant advance within CLS to appear in the same publication as articles written on topics as varied as 3D modelling or database ontology. This paper will nevertheless make an effort to marshal a significant proportion of the field's research output into an historical narrative which is capable of encompassing developments underway in the field since the early sixties. It is in the eighties and nineties that we begin to see previously regnant methods consistently outperformed by multivariate approaches in which ~100 of the most frequent words (MFWs) in a text are quantified. In the oughts and tens CLS scholars extended these methods further, analysing thousands of words and treating texts more or less in their entirety. In accounting for these three phases, this chapter will emphasise particular works of scholarship which have been instrumental in transforming one epoch into the next. Any chronology which accounts for the discipline's history will be a generalising one and will require the omission or simplification of particular phenomena. Some articles anticipate transformations within the discipline which are later to take place and, as we will also see, CLS scholars are sometimes prone to continuing to use methods which have been shown to be inadequate. The periodisation here proposed allows us to introduce both superstructural and infrastructural causes in considering the history of the discipline. The tendency to focus on isolated formal features in the discipline's early days is symptomatic of its inclination to reverse the death of the author at the hands of figures such as Roland Barthes and Michel Foucault, re-emphasising the individual agency and style of the author. In its early history therefore, CLS manifests an inclination towards more romantic theories of authorship, a tendency which results in the development of methods which assume that all texts written by different authors are differentiable on the basis of parameters which are wholly arbitrary. As we will see, it is not until the success of John Burrows' Delta method that this notion begins to be challenged.

2 Embryonic CLS (1963-1979)

As Jack Grieve notes, there is a long history of mathematics being brought to bear on the study of attributing authorship, reaching back to the nineteenth century (Grieve 2007, 251). However, in his history of the field, David Holmes identifies the first instance of modern stylometry in Frederick Mosteller and David Wallace's attempts to iden-

tify authorship in the twelve pseudonymously written essays and articles in *The Federalist Papers* written by Alexander Hamilton, James Madison and John Jay in the late eighteenth century. Mosteller and Wallace attribute authorship on the basis of similar rates at which function words are deployed in the text, such as prepositions, conjunctions and articles (Holmes 1998, 112). Fred Damerau seems to be the first to account for the use of function words from a theoretical perspective, citing W.J. Paisley's theory of "minor encoding habits". According to Paisley, in turn drawing from theories developed in the field of art history, indices of personal style can be found in minor, but highly common, features of a work. They should not vary significantly between works produced by the same author but should vary significantly between works produced by different authors. In satisfying these criteria, Damerau identifies function words as being most suitable (Damerau 1975, 271-2). Damerau's approach is perfectly logical and given that it is wholly appropriate to reduce most authorship attribution problems to what Burrows refers to as a "closed game", where there is a restricted set of texts and candidate authors, function words seem to be capable of providing promising results. However, this culminates in the assumption that authorship is in and of itself a guarantor of a distinctive or individual style which suffuses the work in its entirety. It is therefore assumed that each text produced by a single author is statistically homogenous and that any given quantity of features identified in a text written by one author will be statistically distinct from the same feature in a text written by another author, under the assumption that this can be confirmed through the use of Mann-Whitney, chi-square, Student's *t* and Fisher tests. As this paper proceeds, we will see that the influence of this assumption is detrimental. Barron Brainerd's work, in its capacity to identify and willingness to test the resilient assumption of intra-authorial heterogeneity, represents an exception and Brainerd is therefore among the first to identify many of the drawbacks associated with the use of the chi-square method when applied to literary texts (Brainerd 1975, 161; 1979, 5-12).

Though there are significant numbers of papers in *Computers and the Humanities*' (*C&H*) and *Literary and Linguistic Computing*'s (*LLC*) early history which abide by sound statistical and methodological practice, such as Paule Sainte-Marie, Pierre Robillard and Paul Bratley's application of principal component analysis (PCA) to 44 MFWs in 30 plays written by Molière (Sainte-Marie, Robillard, Bratley 1973, 136) and Brainerd's application of cluster analysis in order to differentiate novels from romances (Brainerd 1973, 267), many CLS articles until the nineties can be characterised by the arbitrariness of their methodological approaches. Sampling, variable selection and statistical measurements are often adopted and applied without explicit reasoning or reference to previously undertaken studies with-

in which the efficacy of these methods have been validated. Citations are also less common in early articles than they later become, and this has the effect that the precise rationale for any given procedure being carried out is more often assumed than explained. Robert Cluett's analyses of part-of-speech (POS) entities in Restoration-era prose and John Foley's analyses of stress patterns in Beowulf represent another tendency rife at this early stage in CLS history, which takes a heuristic approach to drawing conclusions rather than using proven mathematical techniques (Cluett 1971, 264-8; Foley 1978, 78). These defects can probably be accounted for by bearing in mind the nascency of the field. As M.W.A. Smith notes, at the time of writing in 1987 there was no extant corpus of studies undertaken which had successfully inculcated an understanding of statistical best practice when analysing literary texts and CLS scholars could not benefit from a corpus of articles on which to base their approaches in the same way a would-be CLS scholar could today (Smith 1987, 146). Other constraints which exert a significant influence on the early scholarship include the available infrastructure. Computational memory limits, which would have been a factor in experimental design, go some way also in explaining the methodological focus we see on quantifying the frequencies of a very small number of function words. Computing was also expensive and, prior to the sharing of digital texts via the internet, each researcher would need to build their own corpus (Sainte-Marie, Robillard, Bratley 1973, 131-2; Sula, Hill 2019, 191).

The early polemics which we find in the first issues of *C&H* are illustrative as regards the 'theory wars', a consistent feature of CLS discourse. It is Louis Tonko Milic who initiates this dialogue, both in *A Quantitative Approach to the Style of Jonathan Swift* (1967) and in two articles which argue for the significance and contributions computing may potentially make to the study of literature. Milic's arguments are based on the capacity of computing to alert the critic or analyst to patterns and trends which are not detectable via traditional, qualitative approaches. This is particularly important as, from Milic's perspective, words which are traditionally deployed in the interrogation or analysis of style in literary criticism are vague or impressionistic. Milic partly attributes this to the blurring of the boundary between literary criticism and social theory (Milic 1967, 27-8, 38, 54). In solving this problem, Milic wished to facilitate a synthesis between computation and the creative intuition which has historically predominated within literary criticism rather than automating the latter out of existence (1966, 5). Milic begins from the notion that syntax may provide a deep and unifying structure or promising a starting point for quantitative approaches (1967, 32, 79) and proceeds by dividing words into twenty-four different grammar-types, looking at how the means of these word-types increase or decrease in Swift's writings over time. Milic then carries out close readings

e-ISSN 2724-3923

of these grammar types in their context within the works (1967, 32, 79, 174, 205, 272). It is Emmanuel Mesthene who presents the first sceptical response, arguing that for all the precision and accuracy which computational tools have the potential to introduce, they also bring bias to literary-critical research as computing cannot serve as a neutrally clarifying agent (Mesthene 1969, 2). Bruce A. Beatie cites C.P. Snow's essay "The Two Cultures" (1959), in locating literary studies within a school of thought totally opposed to that of statistics (Beatie 1979, 186-7). Susan Wittig objects to CLS on the basis of a more overt commitment to post-structuralism, which envisions the text as an ineffable system of exchange which resists all forms of hierarchical categorisation (Anderson 1983, 68). This is utterly contrary to the ways in which natural language processing (NLP) and linguistic analysis require us to regard text (Wittig 1977, 211-2). Despite being written more than half a century ago, these three critics broadly anticipate the two opposed positions we now confront in considering CLS' relationship with the broader literary-critical milieu, even to the present day. Milic, on the one hand, emphasises the capacity of computation to allow the critic to exceed their individual point of view and potentially gain access to an hypothesised deep structure, while CLS' detractors mount an overall objection to CLS in principle, refraining from engaging with statistical methods themselves or a history of their application on the basis that empiricism is an inveterately instrumentalised and insufficiently reflexive form of knowledge production. As this chapter continues, we will see that these two positions and the tensions residing within them are crucial to any account of CLS' history.

3 PCA & Proto-Delta (1980-1990)

In the eighties we see John Burrows publish analyses that anticipate the Delta method he would later develop. Burrows begins by focusing on the changing rates at which modal auxiliaries are used in six novels written by Jane Austen (Burrows 1986, 9). Though Burrows argues his approach allows for the treatment of texts in their entirety, against literary criticism's historical tendency to focus on highly specific features of a work, in his focus on modal auxiliaries and how they relate to sentence length, Burrows remains constrained within the framework he aims to supersede (20-3). In his second article in *C&H*, Burrows attempts to quantitatively differentiate three different narrative categories which he identifies as being at work in Austen's novels; dialogue, 'pure narrative' – here meaning the voice of the narrator alone – and 'character narrative', here meaning the voice of the narrator mediated by the thoughts or feelings of a particular character, elsewhere referred to within literary criticism as 'free in-

direct discourse'. Burrows first correlates the frequencies of a list of function words which appear in each of these three categories, then applies a statistical transformation to these correlation coefficients. The aim of this method, PCA, is to reduce the dimensionality of a dataset consisting of a large number of variables. This is achieved by combining these variables into new variables called 'principal components'. Each principal component encompasses a specific amount of the variation within the original data, to the extent that a two-dimensional visualisation is generally sufficient to provide an insight into the data's underlying structure (Binongo, Smith 1999; Joliffe 2004, 1). Burrows applies this method in a series of distinct permutations, firstly separating the three different narrative types by gender then by character, describing each time the clustering patterns which can be observed in relation to the literary-critical discourse surrounding Austen (Burrows 1987, 64-9). In his third article, Burrows applies his method to fifteen other nineteenth-century novelists. As before, Burrows is invested in identifying a unique and individual style for each author and though his graph has no temporal component, he argues that each author's *oeuvre* clusters chronologically and that Austen, George Eliot and Elizabeth Gaskell's relative distance from the other authors justifies reading their styles as individual, erecting a movement away from neo-classical prose styles which otherwise predominated in the late eighteenth and early nineteenth centuries (Burrows 1989, 318; Holmes 1998, 113).

Given their capacity to cluster texts on the basis of authorship, genre and era, function words remain central within CLS and we see a number of studies emerge which continue to demonstrate the efficacy of the method (Burrows 1992, 91-103; Craig 1991, 183-5; 1999, 222-40; Tse, Tweedie, Frischer 1998, 141-6). We also see further interrogations of PCA in and of itself in Binongo and Smith's investigations into its mathematical principles (1999, 445-64). Penelope Gurney's and Lyman W. Gurney's application of PCA to MFWs significantly outperforms attempts to attribute authorship on the basis of vocabulary richness, a statistic which is calculated by dividing the number of unique word types by the number of words in the text overall (Gurney P., Gurney L. 1998, 119-30). This result is replicated by Fiona Tweedie and R. Harald Baayen, who note that even measurements for vocabulary richness which are independent of text length are unsuccessful in discriminating texts on the basis of their authorship (Tweedie, Baayen 1998, 323-50). Attempts to identify a length-independent means of quantifying a text's lexical richness, for the logical reason that a shorter text will have far more unique word-types than a longer one, are a consistent fixture of CLS discourse, as we see in Philippe Thoiron's diversity or entropy-based method (1986) or John Baker's attempts to quantify the pace at which new vocabulary enters a writer's work (1988, 38-9). The centrality of vocabulary

richness to CLS may be attributed to theories of intra-authorial heterogeneity, but also to the measure's relative simplicity and comprehensibility. This is probably also the case for the persistence of measurements based on sentence, word and syllable lengths, which are also plagued by similar issues relating to reproducibility (Aoyama, Constable 1999). Gurney and Gurney recommend incorporating more MFWs into future analyses, computing space allowing (1998, 119-30).

Concurrent with the development of reliable multivariate statistical techniques in CLS, we also see previously regnant methods challenged for their failures to operate reliably. Thomas Merriam, for example, demonstrates the unreliability of 'proportionate pairs', a method used by A.Q. Morton, which assumes that particular pairs of words which exist in a fixed ratio to one another between texts are suggestive of shared authorship. Merriam demonstrates that more than random variation can often be observed in works produced by the same author (Merriam 1989, 252-3) while Michael Hilton and Holmes demonstrate the inadequacy of another method developed by Morton, wherein the incidence of two formal features are plotted on a line graph. The two lines are then superimposed on one another and it is determined that any instances in which these lines deviate from one another are indicative of the intervention of a second author. Hilton and Holmes propose a more statistically rigorous variant of this approach, which incorporates the weighting of particular features, but concludes that even with these improvements, they fail to reliably attribute authorship (Hilton, Holmes 1993, 73-80; Holmes 1998, 114). Smith also publishes a number of articles which challenge the use of chi-square tests, on the basis that they are prone to delivering Type II errors (Smith 1985, 3-10) as well as Morton's correspondence analyses, based on obtaining corresponding values of particular words in particular positions and collocation analyses, which quantify occurrences of a prescribed word either followed or preceded by a second prescribed word (Holmes 1998, 202; Smith 1987, 145-6). Smith goes on to criticise CLS scholars for using methods which are insufficiently rigorous and proposes instead analysing the rates at which the first word in every speech appears per 1,000 words in the works of six Elizabethan-era playwrights. Smith demonstrates his method's capacity to correctly identify John Webster as the most likely candidate of the six to have authored *The Duchess of Malfi* (1614) and Ben Jonson as the most likely to have authored *The Alchemist* (1610). On the basis of the seeming capacity of this method to function, Smith proposes George Wilkins as being the most likely to have authored *Pericles* (1619) (Smith 1988, 34-7). In the late eighties and early nineties, we see studies which continue to draw from discredited approaches such as the chi-square tests (McColly 1987, 174), the visual inspection of visualisations (Anderson, McMaster 1989, 343-5; Irizarry 1993, 88; Philippides 1988, 4), but these in-

creasingly represent the exception. Even in instances in which PCA is not deployed, in favour of more generic visualisation of distances, analyses employ increasing numbers of variables (Greenwood 1992, 44-7; 1993, 216-9; Irizarry 1991, 176-8). While approaches such as these neglect to aggregate the results of these dendrograms or line graphs as one would within the context of bootstrapping, they still represent the movement of CLS towards holistic analyses of text and a heterogenous number of quantitative methods.

Criticism of CLS in this era continues to maintain the inadequacy of scientific methods operationalised within literary criticism. Both Roseanne Potter and W. van Peer argue that literary studies weigh evidence in a way which is qualitatively different to statistics, which by necessity requires overlooking the process-like nature of literary expression (Potter 1988, 94; van Peer 1989, 303). The difficulty in providing an account of these debates is that neither side, whether they happen to be invested in maintaining a strong post-structuralist current within literary criticism or CLS scholars who wish to render literary studies more empirical, are interested in clarifying or examining what the other side is doing. Even though the milieu at this time would seem to be ripe for the contribution of a scholar versed in both the historical of statistical methods and continental philosophy, such a synthesis unfortunately never materialises. Rather, the strawman which roughly equates one to reactionary politics and the other to an incoherent admixture of feminism and relativism, remains rife. We only need to consider Fortier's arguments that post-structuralist approaches to literature have moved beyond 'sense and reason' (Fortier 1991, 193) or Milic's that postmodernism, as manifested within the strain Milic regards as responsible for the death of the author, is nothing more than a mixture of 'victimisation theory', and 'Marxism' (Milic 1991, 394) to identify how much more heat than light has been generated in CLS scholars' engagements with literary theory.

4 Delta, Results and Prospects (2000-2020)

Burrows first presents the Delta method in 2001 in an attempt to move CLS beyond the quantification of authorship from within the context of the closed game, wherein only two or three authors may be presented as probable candidates within an analysis. The Delta method's capacity to incorporate large numbers of authors, Burrows contends, will allow for the development of CLS analyses which do not close off potential avenues of interpretation before the analysis has begun. Burrows' first use of the Delta method begins by identifying 30 MFWs, disambiguating some of his chosen MFWs on the basis of their grammatical function and expressing each MFW's frequency as a percentage of the number of words in the text overall. The dis-

tribution of each word is then normalised, such that each frequency is expressed in terms of the number of standard deviations it resides from the mean. The 'Delta score' is the mean of the differences between each word's normalised frequency. Through the use of this method, Burrows demonstrates that works by John Milton are less dissimilar to one another than they are to the works of twenty-four other seventeenth-century English poets. Burrows tests Delta with 150, 120, 100, 80, 60 and finally 40 MFWs, observing a decrease in attributional accuracy with each decline in quantified MFWs (Burrows 2002, 272-82). In an article published in Blackwell's *Companion to Digital Humanities* (2004) which analyses forty seventeenth- and eighteenth-century poems, Burrows divides his 150 chosen MFWs into three groups based on subjective readings of their function and applies Delta to each of them separately, trying to identify which of the three cohorts could be considered to be more denotative of authorship as compared with genre (Burrows 2004).

Before the use of the Delta method was taken up to a significant extent, Hoover published a number of articles which involved the application of distance measurements to word frequencies, albeit without normalising or relativising them. Hoover compares how rates of successful attribution are changed by altering the number of MFWs, sample size, methods of computing distance, or removing dialogue, pronouns or texts with a first-person narrator from the analysis. Hoover's analyses replicate Burrows' most significant overall finding, that the quantification of more MFWs increases the rate at which a text is successfully attributed and the most frequent bigrams such as 'it is', 'to the' and 'of the' may be even more effective in this regard (Hoover 2001, 421-38; 2002, 157-76; 2003, 261-82). As G. Bruce Schaalje et al. demonstrate, Delta does not quite allow CLS to definitely break from the problem of the closed game. By virtue of the way in which Delta operates, it in fact tends towards the generation of false positives if it is applied as a means of attributing authorship (Schaalje et al. 2011, 71-88). Scholars such as Patrick Juola have suggested a means by which Delta's tendency to do so can be reduced, by introducing a distractor corpus of true negatives, thereby raising the bar of similarity required if a text is to be identified as the most similar to any other (Juola 2015, i100-13). Even if Juola's proposed adjustment is successful, the central problematic remains in place and is in fact implicit in Burrows' initial terms of reference. Delta is thus best conceived as a means of analysing style in relational terms, rather than as a means of settling instances of contentious authorship. Yet it is a peculiarity of the early discourse that Delta's capacity to consider style in this manner is not considered to any significant extent. We see this in the context of two studies undertaken by Hoover. Hoover is firstly reticent to incorporate additional function words into an analysis, on the basis that this will lead to the quantification

of formal features which are within the conscious control of the author (Hoover 2004b). In a second study, Hoover attempts to improve attributional success by removing textual features such as contractions or personal pronouns from the analysis and then applies Delta to one or two texts divided into a number of different parts in order to see if Delta will cluster them with one another. Hoover's methods therefore again attempt to return to smaller-scale qualitative readings which emphasise the decisive impact of specific formal features (Hoover 2004c). The ongoing influence of Paisley's theory of minor encoding habits is the best means of accounting for why it is that the results of Delta analyses are so consistently passed over, despite the efforts of scholars such as David Mannion and Peter Dixon, who dispute Hoover's and others' focus on unconscious formal features in favour of understanding some other features as being consciously deployed (Mannion, Dixon 2004).

Hoover is the first analyst who aims to further optimise Delta by making quantitative adjustments to Burrows' original method. Hoover does so by treating positive and negative *z*-transformed relative frequencies differently, either by focusing on higher values, or squaring and summing positive and negative means in a number of different permutations. None of these approaches are successful in outperforming Delta outright (Hoover 2004a, 477-95) but proposed modifications are still widely applied and compared with one another, as in Holmes and Daniel W. Crofts (Holmes, Crofts 2010, 179-97). Daumantas Stanikūnas, Justina Madravickaitė and Tomas Krilavičius apply a further modification known as Eder's Delta, which applies weights to frequencies in order to moderate the influence of infrequent word-types (Stanikūnas, Madravickaitė, Krilavičius 2015, 1-7). Shlomo Argamon also attempts to improve Delta on mathematical grounds. Argamon (2008) points out that Burrows normalises word distributions by mean and standard deviation, an approach which would only make sense if the word frequencies were distributed normally, but applies a Manhattan distance, which assumes a Laplace distribution. Stefan Evert et al. (2017), in a subsequent publication which systematically assesses Delta's performance against that of its subsequent improvements, confirm that, based on results obtained from both English and German reference corpora, word frequency distributions are better represented by a normal than by a Laplace distribution. Given this instance of statistical error, Argamon proposes three improvements. The first is Linear Delta, which retains Manhattan distance but normalises the relative frequencies according to median and spread. The second is Quadratic Delta, which retains Burrows' method of normalising, but applies the more mathematically sound Euclidean distance to the word frequencies and finally, on the basis of Delta's doubtful assumption that word frequencies are independent, introduces a third adjustment, Rotated

Delta, which performs a whitening transformation on the word frequencies in order to render them independent from one another (Argamon 2008). Despite their greater degree of mathematical legitimacy, however, Argamon's approaches do not outperform classic Delta (Evert et al. 2017, 8; Jannidis et al. 2015). Peter W.H. Smith and W. Aldridge argue that, on the basis of the assumptions which Euclidean distance makes and the fact that its accuracy decreases as dimensionality – i.e. the number of MFWs we apply the distance measurement to – increases (Smith, Aldridge 2011), there may be an upper limit beyond which we should not quantify words when conducting a Delta analysis. Smith and Aldridge propose 200-300 MFWs as this upper limit, though, as Fotis Jannidis et al. argue, this figure is probably quite low and it may be a product of the fact that Smith and Alridge's study was based on an analysis of a corpus of poetic texts (Jannidis et al. 2015). Jacques Savoy's study, which applies Kullback-Leibler divergence, Burrows' Classic Delta and chi-square in a bid to identify the optimal number of differentiators, argues for between 300 and 500 terms (Savoy 2013). In demonstrating that cosine distance outperforms classic Delta, Evert et al. also note distinct behaviours at higher MFW frequency ranks; classic Delta peaks at ~1000-1500 MFWs and thereafter maintains more erratic behaviour, whereas cosine distance plateaus (2017, 14). Jan Rybicki and Maciej Eder not only quantify up to 3000 MFWs but also test particular strata, attempting to identify if Delta's success may be specific to a particular frequency rank. On the basis of the results obtained, Rybicki and Eder recommend quantifying the first 3000 MFWs (Rybicki, Eder 2011). Alexis Antonia, Hugh Craig and Jack Elliott investigate whether larger *n*-grams as opposed to individual words are more likely to correctly attribute authorship and find that the efficacy of the parameter varies from corpus to corpus (Antonia, Craig, Elliott 2014). Antonia, Craig and Elliott's conclusion that the optimal parameters and measures vary between corpora seems to be confirmed by studies such as Enrico Tuccinardi's, who demonstrates that character grams are more suitable in shorter documents (Tuccinardi 2016) and Lisa Pearl, Kristine Lu and Anousheh Haghighi's analysis of idiolect in epistolary literature, which allows for the weighting of some features as being more important than others (Pearl, Lu, Haghighi 2017). These findings culminate in the developing tendency towards the application of a diversity of methods applied to a similarly diverse set of parameters, for example discriminative words, word lengths, character-based frequency analysis, word-length, POS tags, measures for vocabulary richness, to which vector space representation PCA, hierarchical clustering, SVM, random forests, *k*-nearest neighbours, Delta or rolling Delta, the application of Delta to sequential windows of text, may be applied (Gladwin, Lavin, Look 2017; Hou, Jiang 2016; Saccenti, Tenori 2015; Sayoud 2012). Rybicki and Eder (2011) attempt

to generalise Delta's functionality by applying it to other languages, attaining high levels of success in French, German, Hungarian and Italian corpora but poorer results for Latin and Polish. Richard S. Forsyth and Phoenix W.Y. Lam as well as Rybicki and Magda Heydel apply Delta to translated texts, in an attempt to identify whether the stylistic signal of the author or translator predominates. Both find that the signal of the original author is more powerful, but the presence of different translators can be identified by comparing two different translations of the same author's works (Forsyth, Lam 2014; Rybicki, Heydel 2013, 708-17). Through the use of bootstrap consensus trees and network analysis, which involve the representation of texts and the relationships between them as discrete entities (Eder 2017), Changsoo Lee (2017) demonstrates that the further two languages are apart linguistically, the more likely it is that the translator's writing style will exert itself in comparison to that of the author.

The basic positions we confront in engaging the debates concerning the supposed incompatibility of CLS within literary criticism will by this stage of this paper be familiar and the argument that CLS is both overly generalising and insufficiently reflexive as a form of scholarly inquiry remains the predominant point of attack (Gooding 2013). However, we have not yet considered CLS scholars who have made a virtue of this charge to a certain extent, as in the literary criticism of Franco Moretti. It should be noted that Moretti's major works such as *Atlas of the European Novel* (1998), *Distant Reading* (2005), *Graphs, Maps, Trees* (2007) and *The Bourgeois* (2013) are not substantively computational or statistical in their approaches, but rather use maps, spreadsheets and diagrams in order to illustrate what are often quite traditional literary-critical hypotheses, holding out the possibility that literary criticism might aspire to the ambition and scope of quantitative sociology (Moretti 2007, 4-30; 2012, 67). Moretti's most notorious argument, that the development of industrial capitalism in nineteenth-century Europe (2012, 16-8; 2013, 14-21) paves the way for the emergence of modernist literature is not in and of itself a controversial one; this axiom more or less undergirds a significant amount of literary criticism conducted from a Marxian perspective. Moretti's reception has more to do with what is perceived as his method's apologia for the literary-critical school referred to as the world literary system as it has been developed by Pascale Casanova (Cleary 2006). Criticism of this school has been trenchant from postcolonial scholars such as Emily Apter and Christopher Prendergast on the basis of its tendency towards national chauvinism, imperialist logic and uncritical handling of the relationship between modernisation and the canonisation of literature (Prendergast 2004; Apter 2013, 42-58). However, the publication of Moretti's writings, and responses to them, in pre-eminent venues such as *n+1* and *New Left Review* (Allison et al. 2012; Moretti 2020) has the consequence that

these criticisms have a tendency to assume the shape of criticisms of CLS in general, despite the lack of actual quantification in Moretti's work. That Moretti's far less provocatively post-political analyses of POS tags and word frequencies in the context of the Stanford Lit Lab (Algee-Hewitt, Heuser, Moretti 2015; Allison et al. 2013, 2011) have not been critiqued to the same extent attests to the fact that it is Moretti's more traditional literary-critical work which can be criticised on the basis of its Eurocentricity

Critics who continue to maintain CLS scholars' dependence on reductive or categorical reasoning at this time begin to advocate for more exploratory or interpretative approaches (Escobar 2016, 85; Sinclair 2003) and we might consider Steven Ramsay, Joanna Drucker, Bethany Nowviskie and Jerome McGann symptomatic of this tendency, given their proposals that humanities computing reconfigure itself as a synthesis of theory, statistics and aesthetics. In seeking to locate a common ground between the works of these critics, we might identify their joint rejection of ground truth. The bureaucratic overtones of any reductive striving towards 'accuracy' is eschewed, in favour of a focus on a generative or procedural critical project which may emerge from the transformation of texts, according to the notion that deformance, re-mediation, translation and misprision form crucial parts of the critical enterprise (Ramsay 2011, x; Drucker 2014; Rockwell 2003). The difficulty in considering the work of these critics within the context of CLS is that, even though they may provide novel and engaging philosophical insights, they do not engage to a significant extent with the actuality of statistical approaches and it is as a result impossible, on the basis of their writings, to arrive at practical steps towards the implementation of a provisional or exploratory CLS. There is also a tendency at work in such criticisms to overlook the changing nature of CLS over time.

While, as we have seen, some early CLS scholars may well have had a propensity to overstate the significance of their results, by the 2000s we can see that the promises to reconstruct literary criticism on a foundation more hospitable to scientific rigour in order to exorcise the spectre of post-structuralism have given way to comparisons with endeavours such as sociology, economics or state planning, all of which have long histories of applying statistics in critical and reflective ways. Burrows, for example, asserts that, as it would be an impossibility for a demographer to identify 'pure' instances of the social phenomenon they aim to quantify, whether class, race or gender, the use of spectra or 'fuzzy logic' becomes essential. Burrows' more pragmatic twining of empirical and intuitive analysis undergirded by a growing body of scholarship go a long way towards rejecting the caricature of CLS which sceptics identify as operating within the discipline (Burrows 2018, 725). How much the field of CLS

e-ISSN 2724-3923

can be said to have advanced in this regard can be seen in the work of Taylor Arnold and Lauren Tilton (2019, 4-14) which, in its simultaneous consideration of both Barthes and the functionality of machine learning, takes strides in combining the actual mechanics of computing and theoretical criticism as well as the changing nature of other anti-CLS articles which now circulate. While the criticisms Nan Z. Da presents in *Critical Inquiry* are partly inhibited by their aim to de-legitimise the quantification of literature in general, Da's article still represents a paradigm shift, in that it argues that computational or statistical methods are widely misunderstood or not implemented properly within. Implicit within Da's analysis (2019) then, is the notion that the field could be improved on these bases. Katharine Bode, in a response to Da's article, also notes this distinction, as well as the greater degree of care which needs to be taken in critiquing CLS on the basis of its scientism, given the pivot from objective to greater amounts of subjectivity and uncertainty which are made possible via the modelling of machine learning outputs (Bode 2019). These methods, when they are first operationalised in the nineties, function more or less as black boxes. CLS scholars do not expend significant amount of time examining the actual functionality of the algorithms themselves; the emphasis is more often placed on the algorithm's capacity to identify an optimal number of classes having been given them at the outset. In many ways this is to be expected at an early stage in CLS' history, given that, when it is applied to research questions such as Shakespeare's authorship, there is a relatively constrained set of probable candidates. In this sense, machine learning methods are used in more or less the same way as PCA is, as a means of dimension reduction, rather than grappling with the capacity of the method in and of itself. We might compare this with Ted Underwood's 2014 project, "Understanding Genre in a Collection of a Million Volumes", which aimed to classify page-level data into one of three categories, either prose, poetry or drama. In the course of this project, Underwood demonstrates how the two paradigms of knowledge production held to be in opposition for almost the entirety of CLS' history, the statistical and literary aspects, may be synthesised. Underwood notes that, as literary critics do not understand genre empirically, but rather socially, it makes no sense to enforce a rigid either/or classification, but rather an approach based on a spectrum. Approaches arising from the field of machine learning, with its capacity to score goodness of fit as a figure between zero and one, zero representing total uncertainty and one representing absolute certainty, is uniquely suited. A further safeguard against empirical reductionism is erected by cross-validating the obtained results with human judgement, specifically a group of five readers who were recruited in order to classify literary data page by page through the use of a GUI purpose-built for the project. Through the labour ex-

erted by these readers, who labelled all pages in 414 books, training data for the project was obtained, which was instrumental in the algorithm attaining an agreement rate of 94.5% in identifying prose as opposed to poetry, fiction as opposed to nonfiction and body text as opposed to paratext. The statistical model which was constructed on the basis of this training data was found to be less accurate than human judgment by a margin of just 0.9%. In this way, Underwood's utilisation of machine learning points to the capacity of CLS to utilise ambiguity and shades of difference within an empirical approach (Underwood 2014, 8-12).

5 Conclusion

In providing a history of the development of CLS, this paper has demonstrated that, from an early stage in CLS' history, the frequencies of an undifferentiated selection of high-frequency word types were highly effective in clustering texts together on the basis of their authorship. However, CLS scholars aimed to challenge the predominance of post-structuralist theories of authorship and, as a result, CLS was from its inception subject to robust criticism from a cohort of literary critics who were more invested in theoretical readings and who charged CLS critics as operating within a politically reactionary and reductive form of knowledge production. In response, CLS cleaved from an early stage in its history to organic theories of authorship and a focus on unconsciously deployed formal features within the work. The original discovery regarding the efficacy of highly frequent word types is consequently elided for a significant period of time in favour of focuses on the individual contributions of particular words or word types insofar as these can be re-integrated within a traditional or qualitative literary-critical reading. This remains the case even after Burrows develops the Delta method on which subsequent CLS scholars develop improvements; these analyses are noteworthy for their focus on particular words and apparent reluctance to move into higher and higher frequency strata. Yet again this did not change until scholars such as Maciej Eder and Jan Rybicki enact a sequence of benchmark analyses which make the superiority of quantifying thousands of MFWs irrefutable, as well as the development of highly effective unsupervised machine learning techniques optimised for large datasets with thousands of parameters, within which manual intervention would become impractical or inefficient. The development of machine learning represents a significant riposte to the most well-worn arguments against CLS and will no doubt have a significant role to play in the development of the field in the future.

Bibliography

Algee-Hewitt, M.; Heuser, R.; Moretti, F. (2015). "On Paragraphs. Scale, Themes, and Narrative Form". *Stanford Lit Lab*, Stanford. https://litlab.stanford.edu/LiteraryLabPamphlet10.pdf.

Allison, S.; Gemma, M.; Heuser, R.; Moretti, F.; Tevel, A.; Yamboliev, I. (2013). "Style at the Scale of the Sentence". *Stanford Lit Lab*, Stanford. https://litlab.stanford.edu/LiteraryLabPamphlet5.pdf.

Allison, S.; Heuser, R.; Jockers, M.; Moretti, F.; Witmore, M. (2011). "Quantitative Formalism. An Experiment". *Stanford Lit Lab*, Stanford. https://litlab.stanford.edu/LiteraryLabPamphlet1.pdf.

Allison, S.; Heuser, R.; Jockers, M.; Moretti, F.; Witmore, M. (2012). "Quantitative Formalism". *n+1*, (13). https://nplusonemag.com/issue-13/essays/quantitative-formalism-an-experiment/.

Anderson, C.W.; McMaster, G.E. (1989). "Quantification of Rewriting by the Brothers Grimm". *Computers and the Humanities*, 23(4/5), 341-6. https://doi.org/10.1007/bf02176639.

Anderson, P. (1983). *In the Tracks of Historical Materialism. The Wellek Library Lectures*. London: Verso.

Antonia, A.; Craig, H.; Elliott, J. (2014). "Language Chunking, Data Sparseness, and the Value of a Long Marker List". *Literary and Linguistic Computing*, 29(2), 147-63. https://doi.org/10.1093/llc/fqt028.

Aoyama, H.; Constable, J. (1999). "Word Length Frequency and Distribution in English: Part I. Prose". *Literary and Linguistic Computing*, 14(3), 339-58. https://doi.org/10.1093/llc/14.3.339.

Apter, E. (2013). *Against World Literature. On the Politics of Untranslatability*. London: Verso.

Argamon, S. (2008). "Interpreting Burrows's Delta". *Literary and Linguistic Computing*, 23(2), 131-47. https://doi.org/10.1093/llc/fqn003.

Arnold, T.; Tilton, L. (2019). "Distant Viewing". *Digital Scholarship in the Humanities*, 34, Suppl. 1, i3-i16. https://doi.org/10.1093/digitalsh/fqz013.

Baayen, H.; Van Halteren, H.; Tweedie, F. (1996). "Outside the Cave of Shadows". *Literary and Linguistic Computing*, 11(3), 121-31. https://doi.org/10.1093/llc/11.3.121.

Baker, J.C. (1988). "Pace". *Literary and Linguistic Computing*, 3(1), 36-9. https://doi.org/10.1093/llc/3.1.36.

Beatie, B.A. (1979). "Measurement and the Study of Literature". *Computers and the Humanities*, 13(3), 185-94. https://doi.org/10.1007/bf02395096.

Binongo, J.N.G.; Smith, M.W.A. (1999). "The Application of Principal Component Analysis to Stylometry". *Literary and Linguistic Computing*, 14(4), 445-65. https://doi.org/10.1093/llc/14.4.445.

Bode, K. (2019). "Computational Literary Studies: A Critical Inquiry Online Forum". https://criticinq.wordpress.com/2019/03/31/computational-literary-studies-a-critical-inquiry-online-forum/.

Brainerd, B. (1973). "On the Distinction Between a Novel and a Romance". *Computers and the Humanities*, 7(5), 259-70. https://doi.org/10.1007/bf02395426.

Brainerd, B. (1975). "Statistical Analysis of Lexical Data Using Chi-squared and Related Distributions". *Computers and the Humanities*, 9(4), 161-78. https://doi.org/10.1007/bf02402331.

Brainerd, B. (1979). "Pronouns and Genre in Shakespeare's Drama". *Computers and the Humanities*, 13(1), 3-16. https://doi.org/10.1007/bf02744988.
Burrows, J. (2002). "'Delta'". *Literary and Linguistic Computing*, 17(3), 267-87.
Burrows, J. (2018). "Rho-grams and Rho-sets". *Digital Scholarship in the Humanities*, 33(4), 724-47. https://doi.org/10.1093/llc/17.3.267.
Burrows, J.F. (1986). "Modal Verbs and Moral Principles". *Literary and Linguistic Computing*, 1(1), 9-23. https://doi.org/10.1093/llc/1.1.9.
Burrows, J.F. (1987). "Word-Patterns and Story-Shapes". *Literary and Linguistic Computing*, 2(2), 61-70. https://doi.org/10.1093/llc/2.2.61.
Burrows, J.F. (1989). "'An Ocean Where Each Kind...'". *Computers and the Humanities*, 23(4/5), 309-21. https://doi.org/10.1007/BF02176636.
Burrows, J.F. (1992). "Not Unless You Ask Nicely". *Literary and Linguistic Computing*, 7(2), 91-109. https://doi.org/10.1093/llc/7.2.91.
Burrows, J.F. (2004). "Textual Analysis". Schreibman, S.; Siemens, R.; Unsworth, J. (eds), *A Companion to Digital Humanities*. Oxford: Blackwell. https://doi.org/10.1002/9780470999875.ch23.
Cleary, J. (2006). "The World Literary System". *Field Day Review*, 2, 196-219.
Cluett, R. (1971). "Style, Precept, Personality". *Computers and the Humanities*, 5(5), 257-77. https://doi.org/10.1007/bf02402207.
Craig, D.H. (1991). "Plural Pronouns in Roman Plays by Shakespeare and Jonson". *Literary and Linguistic Computing*, 6(3), 180-6. https://doi.org/10.1093/llc/6.3.180.
Craig, H. (1999). "Contrast and Change in the Idiolects of Ben Jonson". *Computers and the Humanities*, 33(3), 221-40. https://doi.org/10.1023/A:1002032032618.
Da, N.Z. (2019). "The Computational Case against Computational Literary Studies". *Critical Inquiry*, 45(3), 601-39. https://doi.org/10.1086/702594.
Damerau, F.J. (1975). "The Use of Function Word Frequencies as Indicators of Style". *Computers and the Humanities*, 9(6), 271-80. https://doi.org/10.1007/bf02396290.
Dixon, P.; Mannion, D. (1993). "Goldsmith's Periodical Essays: A Statistical Analysis of Eleven Doubtful Cases". *Literary and Linguistic Computing*, 8(1), 1-19. https://doi.org/10.1093/llc/8.1.1.
Drucker, J. (2014). *Graphesis. Visual Forms of Knowledge Production*. Cambridge: Harvard University Press.
Eder, M. (2017). "Visualization in Stylometry: Cluster Analysis Using Networks". *Digital Scholarship in the Humanities*, 32(1), 50-64. https://doi.org/10.1093/llc/fqv061.
Escobar, M.V. (2016). "The Essay/ontology Workflow, Challenges in Combining Formal and Interpretive Methods". *Literary and Linguistic Computing*, 31(1), 84-94. https://doi.org/10.1093/llc/fqu071.
Evert, S.; Proisl, T.; Jannidis, F.; Reger, I.; Pielström, S.; Schöch, C.; Vitt, T. (2017). "Understanding and Explaining Delta Measures for Authorship Attribution". *Digital Scholarship in the Humanities*, 32(2), ii4-ii16. https://doi.org/10.1093/llc/fqx023.
Foley, J.M. (1978). "A Computer Analysis of Metrical Patterns in Beowulf". *Computers and the Humanities*, 12(1/2), 71-80. https://doi.org/10.1007/bf02392918.
Forsyth, R.S. (1999). "Stylochronometry with Substrings, or: A Poet Young and Old". *Literary and Linguistic Computing*, 14(4), 467-77. https://doi.org/10.1093/llc/14.4.467.

Forsyth, R.S.; Holmes, D.I.; Tse, E.K. (1999). "Cicero, Sigonio, and Burrows: Investigating the Authenticity of the Consolatio". *Literary and Linguistic Computing*, 14(3), 375-400. https://doi.org/10.1093/llc/14.3.375.

Forsyth, R.S.; Lam, P.W.Y. (2014). "Found in Translation: To What Extent is Authorial Discriminability Preserved by Translators?". *Literary and Linguistic Computing*, 29(2), 199-217. https://doi.org/10.1093/llc/fqt018.

Fortier, P.A. (1991). "Theory, Methods and Applications: Some Examples in French Literature". *Literary and Linguistic Computing*, 6(3), 192-6. https://doi.org/10.1093/llc/6.3.192.

Gladwin, A.A.G.; Lavin, M.J.; Look, D.M. (2017). "Stylometry and collaborative authorship: Eddy, Lovecraft, and 'The Loved Dead'". *Digital Scholarship in the Humanities*, 32(1), 123-40. https://doi.org/10.1093/llc/fqv026.

Gooding, P. (2013). "Mass Digitization and the Garbage Dump: The Conflicting Needs of Quantitative and Qualitative Methods". *Literary and Linguistic Computing*, 28(3), 425-31. https://doi.org/10.1093/llc/fqs054.

Greenwood, H.H. (1992). "St Paul Revisited–a Computational Result". *Literary and Linguistic Computing*, 7(1), 43-47. https://doi.org/10.1093/llc/7.1.43.

Greenwood, H.H. (1993). "St Paul Revisited–Word Clusters in Multidimensional Space". *Literary and Linguistic Computing*, 8(4), 211-19. https://doi.org/10.1093/llc/8.4.211.

Grieve, J. (2007). "Quantitative Authorship Attribution: An Evaluation of Techniques". *Literary and Linguistic Computing*, 22(3), 251-70. https://doi.org/10.1093/llc/fqm020.

Gurney, P.J.; Gurney, L.W. (1998). "Authorship Attribution of the Scriptores Historiae Augustae". *Literary and Linguistic Computing*, 13(3), 119-31. https://doi.org/10.1093/llc/13.3.119.

Hilton, M.L.; Holmes, D.I. (1993). "An Assessment of Cumulative Sum Charts for Authorship Attribution". *Literary and Linguistic Computing*, 8(2), 73-80. https://doi.org/10.1093/llc/8.2.73.

Holmes, D.I. (1998). "The Evolution of Stylometry in Humanities Scholarship". *Literary and Linguistic Computing*, 13(3), 111-17. https://doi.org/10.1093/llc/13.3.111.

Holmes, D.I.; Crofts, D.W. (2010). "The Diary of a Public Man: A Case Study in Traditional and Non-traditional Authorship Attribution". *Literary and Linguistic Computing*, 25(2), 179-97. https://doi.org/10.1093/llc/fqq005.

Hoover, D.L. (2001). "Statistical Stylistics and Authorship Attribution: An Empirical Investigation". *Literary and Linguistic Computing*, 16(4), 421-44. https://doi.org/10.1093/llc/16.4.421.

Hoover, D.L. (2002). "Frequent Word Sequences and Statistical Stylistics". *Literary and Linguistic Computing*, 17(2), 157-80. https://doi.org/10.1093/llc/17.2.157.

Hoover, D.L. (2003). "Frequent Collocations and Authorial Style". *Literary and Linguistic Computing*, 18(3), 261-86. https://doi.org/10.1093/llc/18.3.261.

Hoover, D.L. (2004a). "Delta Prime?". *Literary and Linguistic Computing*, 19(4), 477-95. https://doi.org/10.1093/llc/19.4.477.

Hoover, D.L. (2004b). "Multivariate Analysis and the Study of Style Variation". *Literary and Linguistic Computing*, 18(3), 341-60. https://doi.org/10.1093/llc/18.4.341.

Hoover, D.L. (2004c). "Testing Burrows's Delta". *Literary and Linguistic Computing*, 19(4), 453-75. https://doi.org/10.1093/llc/19.4.453.
Hou, R.; Jiang, M. (2016). "Analysis on Chinese quantitative stylistic features based on text mining". *Digital Scholarship in the Humanities*, 31(2), 357-67. https://doi.org/10.1093/llc/fqu067.
Irizarry, E. (1991). "One Writer, Two Authors: Resolving the Polemic of Latin America's First Published Novel". *Literary and Linguistic Computing*, 6(3), 175-9. https://doi.org/10.1093/llc/6.3.175.
Irizarry, E. (1993). "The Two Authors of Columbus' Diary". *Computers and the Humanities*, 27(2), 85-92. https://doi.org/10.1007/bf01830301.
Jannidis, F.; Pielström, S; Schöch, C.; Vitt, T. (2015). "Improving Burrows' Delta – An Empirical Evaluation of Text Distance Measures". *Digital Humanities Conference 2015*. https://www.researchgate.net/publication/280086768_Improving_Burrows%27_Delta_-_An_empirical_evaluation_of_text_distance_measures.
Joliffe, I.T. (2004). *Principal Components Analysis*. New York: Springer.
Juola, P. (2015). "The Rowling Case: A Proposed Standard Analytic Protocol for Authorship Questions". *Digital Scholarship in the Humanities*, 30, Suppl. 1, i100-i13. https://doi.org/10.1093/llc/fqv040.
Lee, C. (2017). "Do Language Combinations Affect Translators' Stylistic Visibility in Translated Texts?" *Digital Scholarship in the Humanities*, 33(3), 592-603. https://doi.org/10.1093/llc/fqx056.
Mannion, D.; Dixon, P. (2004). "Sentence-length and Authorship Attribution: the Case of Oliver Goldsmith". *Literary and Linguistic Computing*, 19(4), 497-508. https://doi.org/10.1093/llc/19.4.497.
McColly, W.B. (1987). "Style and Structure in the Middle English Poem 'Cleanness'". *Computers and the Humanities*, 21(3), 169-176. https://doi.org/10.1007/bf02252793.
Merriam, T. (1989). "An Experiment with the Federalist Papers". *Computers and the Humanities*, 23(3), 251-4. https://doi.org/10.1007/bf00056147.
Mesthene, E.G. (1969). "Technology and Humanistic Values". *Computers and the Humanities*, 4(1), 1-10. https://doi.org/10.1007/bf02393443.
Milic, L. (1991). "Progress in Stylistics: Theory, Statistics, Computers". *Computers and the Humanities*, 25(6), 393-400. https://doi.org/10.1007/bf00141189.
Milic, L.T. (1966). "The Next Step". *Computers and the Humanities*, 1(1), 3-6. https://doi.org/10.1007/bf00188010.
Milic, L.T. (1967). *A Quantitative Approach to the Style of Jonathan Swift*. Berlin: Mouton & Co.
Moretti, F. (1998). *Atlas of the European Novel 1800-1900*. London: Verso.
Moretti, F. (2007). *Graphs, Maps, Trees. Abstract Models for a Literary History*. London: Verso.
Moretti, F. (2012). *Distant Reading*. London: Verso.
Moretti, F. (2013). *The Bourgeois. Between History and Literature*. London: Verso.
Moretti, F. (2020). "The Roads to Rome". *New Left Review*, 2(124), 125-36.
Pearl, L.; Lu, K.; Haghighi, A. (2017). "The Character in the Letter: Epistolary attribution in Samuel Richardson's *Clarissa*". *Digital Scholarship in the Humanities*, 32(2), 355-76. https://doi.org/10.1093/llc/fqw007.
Philippides, D.M.L. (1988). "Literary Detection in the Erotokritos and The Sacrifice of Abraham". *Literary and Linguistic Computing*, 3(1), 1-11. https://doi.org/10.1093/llc/3.1.1.

Potter, R.G. (1988). "Literary Criticism and Literary Computing: The Difficulties of a Synthesis". *Computers and the Humanities*, 22(2), 91-7. https://doi.org/10.1007/bf00057648.

Prendergast, C. (2004). "The World Republic of Letters". Prendergast, C. (ed.), *Debating World Literature*. London: Verso, 1-25.

Ramsay, S. (2011). *Reading Machines*. *Towards an Algorithmic Criticism*. Champaign: University of Illinois Press.

Rockwell, G. (2003). "What is Text Analysis, Really?". *Literary and Linguistic Computing*, 18(2), 209-19. https://doi.org/10.1093/llc/18.2.209.

Rybicki, J.; Eder, M. (2011). "Deeper Delta Across Genres and Languages: Do We Really Need the Most Frequent Words?" *Literary and Linguistic Computing*, 26(3), 315-21. https://doi.org/10.1093/llc/fqr031.

Rybicki, J.; Heydel, M. (2013). "The Stylistics and Stylometry of Collaborative Translation: Woolf's *Night and Day* in Polish". *Literary and Linguistic Computing*, 28(4), 708-17. https://doi.org/10.1093/llc/fqt027.

Saccenti, E.; Tenori, L. (2015). "Multivariate Modeling of the Collaboration Between Luigi Illica and Giuseppe Giacosa for the Librettos of Three Operas by Giacomo Puccini". *Digital Scholarship in the Humanities*, 30(3), 405-22. https://doi.org/10.1093/llc/fqu006.

Sainte-Marie, P.; Robillard, P.; Bratley, P. (1973). "An Application of Principal Component Analysis to the Works of Molière". *Computers and the Humanities*, 7(3), 131-7. https://doi.org/10.1007/bf02403851.

Savoy, J. (2013). "Comparative Evaluation of Term Selection Functions for Authorship Attribution". *Digital Scholarship in the Humanities*, 30(2), 246-61. https://doi.org/10.1093/llc/fqt047.

Sayoud, H. (2012). "Author Discrimination Between the Holy Quran and Prophet's Statements". *Literary and Linguistic Computing*, 27(4), 427-44. https://doi.org/10.1093/llc/fqs014.

Schaalje, G.B.; Fields, P.J.; Roper, M.; Snow, G.L. (2011). "Extended Nearest Shrunken Centroid Classification: A New Method for Open-Set Authorship Attribution of Texts of Varying Sizes". *Literary and Linguistic Computing*, 26(1), 71-88. https://doi.org/10.1093/llc/fqq029.

Sinclair, S. (2003). "Computer-Assisted Reading: Reconceiving Text Analysis". *Literary and Linguistic Computing*, 18(2), 175-84. https://doi.org/10.1093/llc/18.2.175.

Smith, M.W.A. (1985). "An Investigation of Morton's Method to Distinguish Elizabethan Playwrights". *Computers and the Humanities*, 19(1), 3-21. https://doi.org/10.1007/bf02259614.

Smith, M.W.A. (1986). "A Critical Review of Word-links as a Method for Investigating Shakespearean Chronology and Authorship". *Literary and Linguistic Computing*, 1(4), 202-6. https://doi.org/10.1093/llc/1.4.202.

Smith, M.W.A. (1987). "Hapax Legomena in Prescribed Positions: An Investigation of Recent Proposals to Resolve Problems of Authorship". *Literary and Linguistic Computing*, 2(3), 145-52. https://doi.org/10.1093/llc/2.3.145.

Smith, M.W.A. (1988). "The Authorship of Acts I and II of *Pericles*: A New Approach Using First Words of Speeches". *Computers and the Humanities*, 22(1), 23-41. https://doi.org/10.1007/bf00056347.

Smith, P.W.H.; Aldridge, W. (2011). "Improving Authorship Attribution: Optimizing Burrows' Delta Method*". *Journal of Quantitative Linguistics*, 18(1), 63-88. https://doi.org/10.1080/09296174.2011.533591.

Stanikūnas, D.; Mandravickaitė, J.; Krilavičius, T. (2015). "Comparison of Distance and Similarity Measures for Stylometric Analysis of Lithuanian Texts". *International Conference for Young Researchers in Informatics, Mathematics and Engineering*. http://ceur-ws.org/Vol-1852/p01.pdf.

Sula, C.A.; Hill, H.V. (2019). "The Early History of Digital Humanities: an Analysis of Computers and the Humanities (1966-2004) and Literary and Linguistic Computing (1986-2004)". *Digital Scholarship in the Humanities*, 34, Suppl. 1, i190-i206. https://doi.org/10.1093/llc/fqz072.

Thoiron, P. (1986). "Diversity Index and Entropy as Measures of Lexical Richness". *Computers and the Humanities*, 20(3), 197-202. https://doi.org/10.1007/bf02404461.

Tse, E.K.; Tweedie, F.J.; Frischer, B.D. (1998). "Unravelling the Purple Thread: Function Word Variability and the *Scriptores Historiae Augustae*". *Literary and Linguistic Computing*, 13(3), 141-9. https://doi.org/10.1093/llc/13.3.141.

Tuccinardi, E. (2016). "An Application of a Profile-Based Method for Authorship Verification: Investigating the Authenticity of Pliny the Younger's Letter to Trajan Concerning the Christians". *Digital Scholarship in the Humanities*, 32(2), 435-47. https://doi.org/10.1093/llc/fqw001.

Tweedie, F.J.; Baayen, R.H. (1998). "How Variable May a Constant be? Measures of Lexical Richness in Perspective". *Computers and the Humanities*, 32(5), 323-52. https://doi.org/10.1023/A:1001749303137.

Underwood, T. (2014). "Understanding Genre in a Collection of a Million Volumes. Interim Report". https://doi.org/10.6084/m9.figshare.1281251.

van Peer, W. (1989). "Quantitative Studies of Literature. A Critique and an Outlook". *Computers and the Humanities*, 23(4/5), 301-7. https://doi.org/10.1007/bf02176635.

Wittig, S. (1977). "The Computer and the Concept of Text". *Computers and the Humanities*, 11(4), 211-15. https://doi.org/10.1007/bf02396857.

e-ISSN 2724-3923

magazén
Vol. 1 – Num. 2 – December 2020

Digital Public History Inside and Outside the Box

Enrica Salvatori
Università di Pisa, Italia

Abstract The relationship between history and information technology is long and troubled. Many projects have opened and closed, many pioneering initiatives have achieved success and attention, but often they have not 'taught' or left a fruitful legacy. These experiments, studies and conferences have built a rich basis for the relationship between the disciplines, but the difficult connection needs to be further explored. Although no digital history research centres exist in Italy nowadays, it is possible to look at several Italian projects in order to discuss the strange position of the digital issue within public history conferences, the place of history in the large digital humanities environment and finally the definition itself of digital history.

Keywords Digital history. Digital humanities. Public history. History. Computer science.

Edizioni Ca'Foscari

Peer review

Submitted	2020-07-07
Accepted	2020-09-16
Published	2020-12-22

Open access

Citation Salvatori, E. (2020). "Digital Public History Inside and Outside the Box". *magazén*, 1(2), 203-222.

DOI 10.30687/mag/2724-3923/2020/02/003

1 Introduction

There is a long-standing relationship between historians and information technology. Historians were among the first humanists to welcome the emergence of new analysis tools developed by ITC. Several pioneering projects, studies and conferences have shaped a difficult and often unsatisfactory relationship between the two areas.[1] Instead of retracing this long and intricate relationship, we will focus on recent Italian meetings and current Italian digital humanities projects. The aim of this essay will be to present and extrapolate the role of digital analysis in the historical Italian researchers' work. [2]

The meetings taken into consideration are the 2018-2019 national conferences of the Italian Association of Public History (AIPH)[3] and the 2018-2020 national conferences of the Italian Association for Digital Humanities and Culture (AIUCD)[4] on the basis of the respective abstract books. The selected abstracts from the 2018 first conference of the Italian Association of Medieval Historians (SISMED)[5] were also taken into account. SISMED is the only group among the national Coordination of Historical Societies[6] that chose this form of communication for their recurring meeting. The evolution of historical research can be inferred from the proceedings of these institutional meetings, which were organised by national historical associations. The accepted abstracts are concise, structured descriptions with keywords, but not yet filtered by the long process of peer review of the scholarly journals. The collected data has been compared with the Italian digital humanities projects highlighted on the AIUCD website[7] and with the essays of a recent book dedicated to Digital History (Paci 2019).

1 Gil 2015, 161-78; 2019, 177-81. From this point of view, the autobiography of Manfred Thaller is extremely interesting (Thaller 2017, 7-109).

2 All translations, if not otherwise stated, are made by the Author.

3 https://aiph.hypotheses.org/.

4 http://www.aiucd.it/.

5 http://www.sismed.eu/.

6 http://www.sissco.it/articoli/componenti-del-coordinamento-delle-societa-storiche/. The coordination includes the Central Council for Historical Studies, the Italian Association of Public History (AIPH), the University Council for the History of Christianity and Churches (CUSCC), the University Council for Greek and Roman History (CUSGR), the Society of Italian Economic Historians (SISE), the Italian Society of Historians Women (SIS), the Italian Society for the History of the Modern Age (SISEM), the Italian International History Societies (SISI), the Italian Association of Medieval Historians (SISMED), the Italian Society for the Study of Contemporary History (SISSCO).

7 http://www.aiucd.it/progetti/.

magazén
1, 2, 2020, 203-222 | e-ISSN 2724-3923

2 Digital Humanities in AIPH and in SISMED

Although public history is a field of historical research that has long been widespread in several nations, it is only recently that this 'discipline' has been galvanised in Italy with the creation – unprecedented in Europe – of the AIPH and the organisation of three highly attended national conferences.[8] This subject is peculiar and interesting for our purposes, because it is highly diachronic and transdisciplinary, given that it has no chronological limits and collects initiatives promoted by very different professional actors, united by the red thread of history: historians, documentary filmmakers, journalists, archivists, museologists, librarians, photographers, cultural operators, re-enactors etc. In short, in the public history field an extremely wide concept of history is used by a large range of people who 'make history', in a way that sometimes looks too dispersive, but by bucking the heavy disciplinary specialisation of the academic world.[9] However, concerning our analysis – the importance of digital history for the production of knowledge by historians –, the AIPH books offer different qualitative and quantitative data in comparison to what is found in a traditional miscellaneous volume of ancient, medieval, modernist or contemporary historians. In fact, our society values the promotion of participatory projects and the involvement of the public. This makes digital tools and methods often indispensable and relevant for the public historians, compared to a more traditional research practice in the academic environment. Unfortunately, a comparative analysis of the transversal historiographical production of the academic Italian world is greatly hampered by the variety and the quantity of the sector's scholarly journals. As mentioned, there has been only one book of abstracts published by a traditional historians' national association. The analysis of this unique piece, the collection of SISMED documents composed just after the first national conference (SISMED 2019),[10] provides the following data: out of 140 interventions, distributed in 48 panels, a significant use of typical digital humanities tools and methods has been found only in

8 Ravenna 2017, Pisa 2018, Santa Maria Capua a Vetere-Caserta 2019. The fourth (Venezia-Mestre 2020) was cancelled because of the pandemic COVID-19. On the Public History in Italy and in the international framework, see Noiret 2009, 275-327; Noiret 2011, 10-35; Cauvin 2018, 3-26.

9 The Italian Ministry of Education, University and Research (MIUR) provides up to 77 different sectors for the area 10 *Sciences of antiquity and philological-literary and historical-artistic sciences*, even if with exquisitely literary and linguistic disciplines; the area 11 *Historical, philosophical, pedagogical and psychological sciences* has 34 sectors: `https://www.miur.it/UserFiles/116.htm`.

10 SISMED 2019. This book consists of a simple juxtaposition of the original texts by the editors, without any editorial homogenization; therefore, there are considerable differences in the length and in the structure of the texts, which limits the possibility of their comparative analysis.

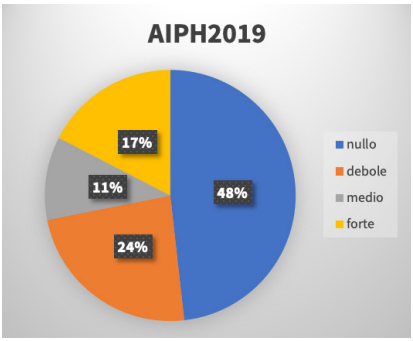

Figure 1
Comparison of the 'weight' of the 'digital' in papers analysed during AIPH 2019

2 panels (4.2%) and 6 papers (4.2%); in no case was the digital issue a central topic, not even in order to address methodical issues.

Let us now look at the data obtained from AIPH 2018 and AIPH 2019 (Salvatori, Privitera 2018; Santarelli 2019). A first analysis concerned the importance of the 'digital' in the papers according to four levels: unsatisfactory, weak, medium and strong. See figure 1 for the results (translation: *nullo* = zero or unsatisfactory, *debole* = weak, *medio* = medium and *forte* = strong) **[fig. 1]**.

The number of proposals without connection to the digital world and its tools has strongly decreased and the attention to the new technologies has deepened at the same time; the second remarkable feature is that the 'medium' and 'strong' values largely exceed those expressed by the SISMED's book for both years.

After carefully analysing which tools and methods of digital humanities are being privileged in the domain of public history, though considering only presentations for the medium employed and the strong proposal, the results of this inquiry are shown in figure 2 **[fig. 2]**.

The most interesting areas of the digital world are the web and social media. Public History activities take place mainly in the 'real' world and in 'material' projects, but this field undoubtedly received a consistent boost from the web, finding new and extremely effective tools for interacting with its own various audiences. I do not refer to a simple presence on the web – as a showcase –, nor to a new digital shape given to the traditional dissemination: the use of the web and social media in a good public history project needs a well balanced and complementary use of different media organised around the web interface; this interface turns into a complex construction, with qualified maintenance duties, when it becomes a gateway for collaborative projects for digital source collections (by crowdsourcing). On the web the public historian knows not only how to organise contents and manage collections but has to turn him/herself into a 'manager of participation', skilled both with a good basic competence

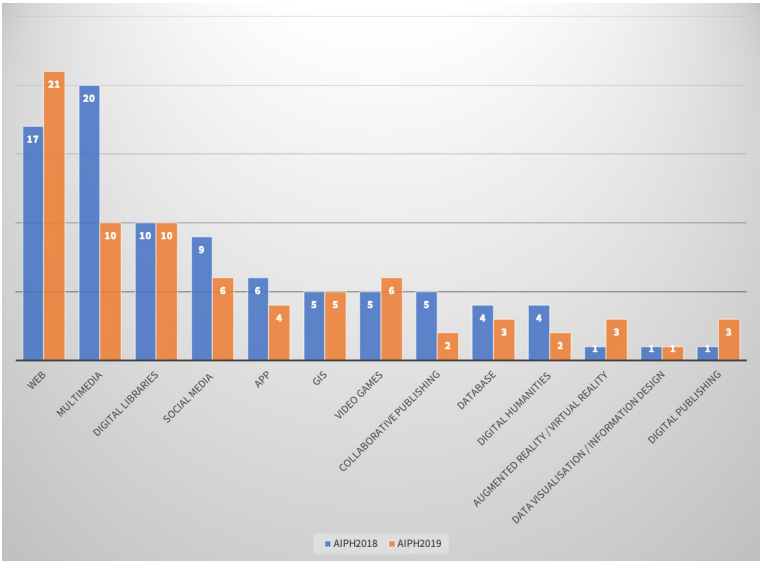

Figure 2 Tools and methods of digital humanities in Italian public history congresses

on different digital tools, and with a deep understanding of communicative methods and languages.

This situation clearly emerged in a study I conducted with the students of the master's degree program in Digital Humanities in Pisa from 2017-2018. In collaboration with AIPH, in September 2017 I promoted a spontaneous collection of best Public History practices in order to better define the extremely manifold panorama of practices that emerged from the 1st national conference in Ravenna (June 2017) and to propose guidelines for the promotion and the implementation of Public History initiatives.[11] The projects were examined and discussed within the Digital Public History course I held, through the development and the use of an evaluation grid based on authorship, fairness, transparency, methodological validity, participation and role of the public historian. One of the aspects that the students have most evaluated and discussed was the presence of the digital in Public History practices. The other aspect evaluated was the possibility to distinguish Digital History and Digital Public History.

11 The practices were collected by a call spread through the AIPH website and on the main social networks (Facebook and Twitter). The collection was made through a Google form that asked compilers to provide, besides the description of the project, details on the sources used, the relationship with the public, the main medium, the nature of the promoters. See the report in Salvatori 2018.

About the first point, the projects were really heterogeneous, starting from a superficial or immature presence on the web as meta-medium to an aware and highly specialised use of it. The latter level concerned activities that combined the creation of the project site – in order to publish historical sources with metadata – with a differentiated use of the social networks to collect the sources with crowdsourcing practices. Equally another aspect that was highly appreciated was the transition from the video lesson or interview (a digital version of the classic conference) to different formats with video-dialogues shared on social media open to comments and interactions.

In short, the initiatives that made a qualitative leap were those that had consciously used digital humanities tools and methods to implement collaborative practices for history, crowdsourcing initiatives, opportunities for dialogue between private memories and institutional archives, up to virtual reality and multisensory paths. In this sense, the digital public historian seems to acquire the role of the "designer of the historical knowledge […] who apply investigation, analysis, imagination and interpretation as 'techniques' to create meaningful media environments suitable for the communities he/she wants to involve".[12]

Going back to figure 2, the second most popular topic, after the methods and tools related to the web, concerns the techniques and the methodologies related to the construction of Digital Libraries (i.e. the organised collection of digital information) (Tammaro 2005). Other tools seem to have a minor role, although the importance of apps, GIS (geographic information system), and video games must be recognised. Video games, however, are not usually built by the public historian, but only analysed for their possible use in Public History practices.[13]

3 History in the AIUCD Digital Humanities

The AIUCD was established in March 2011. Since its first conference, it collected and published the proceedings of its national initiatives, even if in different formats; the recent update of the site, more or less in correspondence with the birth of the official journal *Umanistica Digitale*,[14] has allowed for a better organisation of these mate-

12 Salvatori 2018, see the contribution of Stefano Capezzuto.

13 With the important exception of the gamification project by Fabio Viola: https://www.tuomuseo.it/.

14 https://umanisticadigitale.unibo.it/.

rials and therefore has also facilitated the analysis.[15] For the years 2018-2020 there are two books of abstract and a book of proceedings, whose structure is really comparable.[16]

Given that all the abstracts are related to the complex galaxy of the digital humanities, my research obviously concerned the presence and role of History as a discipline. Once I identified the historical research, I focused on the tools and methods used to carry out the studies. The first problem in labelling the papers was to define the 'historical' projects. I distinguished them from those dealing with other subjects. Such distinction was extremely difficult to make, although I founded it on consistent and solid motivations: on the one hand, digital humanities are, by nature, interdisciplinary and therefore usually create extremely hybrid communities of practice; on the other hand, working in the field of human sciences, digital humanists always exhibit a relationship with history, whether they take care of the edition of a text, build digital libraries of cultural heritage, model an epigraph in 3D, or focus on the analysis of a phenomenon using Social Network Analysis. However, there is an epistemological difference between a 3D model of a Romanesque capital built to be viewed in a museum's app and one that can be explored on GIS about the medieval iconography: both projects have to do with history and need historical skills, but only the second originates from and directly answers to a historical question.

Likewise, if we look at the wide sector of the digital editions of texts, it is obvious that these publications always facilitate historical research, but there is a clear difference between an edition built to study linguistic data and one that highlights the elements of greatest historical relevance. In order to recognise Digital History within the current Italian Digital Humanities, a painful simplification of complexity was nonetheless necessary: I chose identify some macro areas for the historical papers, based on a few general categories, taking into account the main question at the basis of the research. Therefore, the category 'historical disciplines' has been applied only to the abstracts where the proximity to the historical problems was explicit and predominant (History, Art History, History of Literature, History of Science, Oral History, History of Ideas, History of Architecture), providing different labels for contributions dedicated to the field of Archives, Libraries, Bibliography and Artistic-Architec-

15 On *Umanistica Digitale* in the BoA page there are the Books of Abstract of the 2016-2019 meetings and the proceedings of 2013 and 2020 (`https://umanisticadigitale.unibo.it/pages/view/boa`); in the site also the materials of the 2012, 2014 e 2015 meetings are linked (`http://www.aiucd.it/convegno-annuale/`).

16 The main difference is that for AIUCD2020 authors were asked to submit an extended essay after the first selection. These are the books: Spampinato 2018; Allegrezza 2019; Marras et al. 2020.

e-ISSN 2724-3923

tural Heritage. As for the Disciplines of the Text, they were considered both collectively and by distinguishing them among Publishing, Philology, Literature and Linguistics (the digital edition of historical sources has been included in the 'Philology' label).

The outcome is in table 1:[17]

Table 1 Type of content in AIUCD books of abstract

	AIUCD 2018		AIUCD 2019		AIUCD 2020	
abstracts	**52**	**%**	**57**	**%**	**43**	**%**
Digital humanities	2	3,85	15	26,32	2	4,65
Cultural heritage	14	26,92	13	22,81	10	23,26
Publishing	1	1,92	2	3,51	0	0,00
Philology	11	21,15	5	8,77	8	18,60
Philosophy	0	0,00	2	3,51	0	0,00
Geography	1	1,92	0	0,00	0	0,00
Literature	6	11,54	9	15,79	4	9,30
Linguistic	7	13,46	6	10,53	12	27,91
Historical disciplines	12	23,08	8	14,04	7	16,28

In table 2, the presence of History was isolated among the macro sector of the Disciplines of the Text and the remaining branches of the Digital Humanities.

Table 2 Refined type of content in AIUCD books of abstract

	AIUCD 2018		AIUCD 2019		AIUCD 2020	
abstracts	**52**	**%**	**57**	**%**	**43**	**%**
Disciplines of the text	25	48,08	22	38,60	24	55,81
Cultural heritage	14	26,92	13	22,81	10	23,26
Historical disciplines	12	23,08	8	14,04	7	16,28
History	4	7,69	4	7,02	3	6,98
Other	3	5,77	2	3,51	2	4,65

The three meetings, albeit with some understandable oscillations, offer a fairly stable picture of Italian Digital Humanities, where the projects belonging to the Disciplines of the Text and to the management of the digital (digitised or born digital) cultural heritage are dominant. The world of historians in a broad sense promotes an average

17 Please note that an abstract can have from one to three labels and the 'education' label for AIUCD2019 was expunged as the conference was dedicated to *Pedagogy, Teaching, and Research in the Age of Digital Humanities*.

of 15-17% of the projects, a percentage that decreases considerably if we consider History – in a narrower sense – as a discipline different from other specialised sectors, which obviously developed peculiar methodologies, such as History of Architecture or History of Art.

A closer look at the historical proposals allows us to better recognise the digital humanities tools and methods that have been used in research (overall analysis for the three-year period 2018-2019). In evaluating table 3, consider that, even here, I had to simplify by combining the multiple tools developed in these recent years.

Table 3 Digital tools in historical content in AIUCD books of abstract

	History	History of Arts (Music, Visual Arts)	History of Culture (Literature, Science, Ideas)	History of Architecture	Archaeology	TOT
Database	6	2	1	2	1	12
GIS	4				1	5
Web/social	3		2			5
Social network analisys	2		1	1		4
Multimedia	1		1			2
Text encoding	1		1			2
Corpora	1	1				2
Semantic web			1		1	2
3D modeling				2		2
Digital libraries		1	1			2
Distant reading		1				1
Data visualization		1				1

Apart from the anomalous data of archaeology, which is in fact not represented in the AIUCD activity and therefore cannot be assessed from the point of view of its relationship with digital humanities through the considered sources, the data clearly show important the use of databases is for Digital History and, in particular, GIS. The relationship between history and geography has always been obviously very close, but we can surely say that the GIS has provided the historical, archaeological and architectural disciplines with a key tool that can produce a qualitative leap in cataloguing, analysing, comparing, and visualising historical data. GIS is obviously used in many other fields, but in this analysis, it appears as the main tool in historical projects. This data is certainly not new: see, for example, Franc-

esca Bocchi's pioneering research in Bologna.[18] It also represents a useful reference for any plans to extend the training of historians in the field of the digital humanities.

Then, we cannot overlook the data about the web, which essentially concerns the publication of research outcomes for a wider public, with the contextual use of social networks and multimedia communication formats. Here too the data is not surprising and has a logical explanation: for historians the relationship with their 'audience' is fundamental, to a greater extent than in other human sciences. As we have already seen in the analysis on the AIPH books of abstracts, digital history often tends to naturally become digital public history, both because the historical interpretation has a very close and complex link with the public narrative of history,[19] and because the birth of the web and the social networks has created new historiographic practices open to direct participation of the users. Another phenomenon shown in the table is the distance that still divides historians from the tools that have long been developed and used by linguists, philologists and scholars of literature. All digital humanities essays rightly praise father Busa and identify the Computational Linguistics as the starting point of the new course of the human sciences since the Seventies, but the text encoding techniques, the study of the concordances, and the natural language processing tools are still very far from the practice of the historians, as well as the semantic web and the social networks analysis.

This distance is well exemplified by looking at some ongoing projects reported by the AIUCD website, which, in the vast majority of cases, provide the historian with resources to explore a certain phenomenon, but are not promoted by historians.[20]

The *Vespasiano da Bisticci's Letters* edited by Francesca Tomasi[21] is an excellent example of a digital scholarly edition of a fifteenth century Florentine copyist, accompanied by several philological orientation tools (authorities, synoptic table, philological notes, description of the witnesses) and other information. The historical data, however, are offered only through linked data from the corresponding DBpedia entries and no further information on the historical figure, the period or the historical context can be found on the site. To avoid misunderstandings, I evaluate the edition of Vespasiano da Bisticci's *Letters* as one of the best Italian digital humanities projects currently

18 http://www.centrofasoli.unibo.it/nume/italiano/progetto.html.

19 By public narrative of the History we obviously do not mean only the History conveyed and promoted by the institutions, but also that produced by the communities, the groups, the movements.

20 http://www.aiucd.it/progetti/. It would be a useful and interesting service if the digital humanities projects always included a section that would make the reader aware of the use that is being made of the resource itself.

21 http://vespasianodabisticciletters.unibo.it/.

active: I simply recognise that this source is hardly useful for historical research, not made for and nor thought by historians. A similar statement, albeit with the necessary distinctions, can be done – looking at the AIUCD showcase – for *DanteSources*, *Digital Ramusio*, *Epistolario Alcide de Gasperi*, *Idilli di Giacomo Leopardi*, *Petrarchive*, *La dama boba*, *Last Letters* and *PoLet500*.

The remaining projects on the site showcase are collections of resources that could be filed in the macro-category of the 'Cultural Heritage' collection: digital libraries of historical textual or material or hybrid sources, equipped with a rich set of metadata and searchable through a dedicated engine. These projects include the DASI (Digital Archive for the Study of pre-Islamic Arabian Inscriptions), the CPhCl (Catalogus Philologorum Classicorum), the *Digital Library of Family Books* or *The Uffizi Digitization Project* with the 3D collection of the Greco-Roman sculptures of the Uffizi, Palazzo Pitti and the Boboli Gardens museums. These are essentially digital libraries, organised repositories of objects related to the cultural heritage (therefore valuable resources for the historian), normally built without the presence of a historian on the team and – apart from *The Uffizi Digitization Project* – made to answer to linguistic/literary questions. The only exception in this panorama is *Colonizzazioni interne e migrazioni* (Inner colonizations and migrations),[22] a digital history project that collects, catalogues and geolocates the projects of colonisation promoted by the European chancelleries between the 16th and 18th centuries by involving foreign settlers. Also, *Colonizzazioni interne e migrazioni* offers a repertoire of resources, but it gives priority, compared to the direct consultation of primary sources, to materials half-processed or processed to answer historical questions.

4 A Strange Position

At least two sectors – Disciplines of the Text and Digital Libraries – touch on the strange position of Digital History in current historical research. Historians seem to have completely assigned the competence on the digital edition and the treatment of written sources of the Disciplines of the Text and the responsibility for the management of historical collections, corpora, archives and digital libraries to the professionals of the GLAM sector,[23] who do not play an active

22 https://storia.dh.unica.it/colonizzazioninterne/about by Giampaolo Salice. Exceptions also include the *Codice Pelavicino Digital Edition*, edited by myself: an edition in which the text encoding looks with particular attention at historical data and the web interface is designed for the collaboration with the public; it is not examined because it is still incomplete (http://pelavicino.labcd.unipi.it/).

23 Acronym for Galleries, Libraries, Archives and Museums.

role in designing platforms and survey tools. In these areas, History is practically everywhere, but used as parsley in the cooking: it seasons many dishes, but it does not hold a whole one.

This distance appears less wide in the Digital Public History field, where – as we know well by now – digital libraries rank third in the use of digital tools and methods, after the web and the social media; but that kind of projects are usually promoted by public historians who work in the GLAM world. So, the situation remains basically the same. The web – as we know well by now – provides a huge quantity and variety of content. It allows free access to primary and secondary sources that the users can freely collect, associate, annotate and elaborate in a personal or collective (community) reading. When a digital library not only allows for a good access to the sources, but favours their discovery and rediscovery, enhances certain contents, and brings a peculiar investigation path to the attention of the user in order to answer a collective need or because of its historical relevance, then we are in the field of the Digital Public History. It does not matter if the author is an archivist or an historian.

Among the best examples is the project *Cartastorie* of the Museum of the Historical Archive of the Banco di Napoli,[24] born to enhance the enormous heritage kept in the folders of the ancient Neapolitan banks in perfect complementarity with the database and the digital services of the archives themselves: the *Cartastorie*, with its multimedia paths "respectful of the identity and specificities of the Archives" which address "different audiences in different ways by creating for them an experience of wonder and amazement that is not separated from sense and meaning" is a fine example of intermediation of an archival heritage with the aims and methods of the public historian. This need to reason about a new intermediation, after the great disintermediation of the various web contents formats, comes from a world – the GLAM one – made by people who daily and steadily work on the sources of the historians' work.

5 Digital Tools and Methods: An Educational Problem?

My analysis is mirrored in a recent book edited by Deborah Paci on Digital History, an interesting collection of essays that reflects on the 'digital historical culture' nowadays, presenting a panorama of the ongoing research (Paci 2019).

In the book, practices and research of Digital History are in fact distributed over four main sections or thematic areas, the same that emerged from our study: the communication of historical content at

24 http://www.ilcartastorie.it/.

different levels of complexity and participation (web/social/storytelling), the historical-geographical databases (GIS), the archiving and information retrieval practices (our 'cultural heritage' label), and the use of computational linguistics and text processing methods in the historian's job.

The relative weight of the 4 areas are more or less corresponding to that highlighted in our brief excursus, with a clear prevalence of the use of GIS and the methods and tools related to web communication; less featured are the information retrieval techniques and the computational methods for the analysis of the written sources. As regards the digital archives and libraries, the volume highlights that the search interface, the underlying data model, the quality of the digitization of the texts, and the navigation tools themselves are difficult to use by and therefore unsatisfactory for historians (Maxwell 2019); however, from this point of view, historians obviously pay for their self-isolation, because design teams of digital libraries are rarely involved in the building of those platforms and those who actually build them do not have a sufficient knowledge on information retrieval methods. As regards the Disciplines of the Text, the tools and methods for a semi-automatic analysis of a written source are now widespread, as are the techniques for the social network analysis in textual corpora, but these methods are very rarely used in the Italian historiography of the last 10 years.[25] Likewise, the evolution of the methods to publish a good digital edition of historical sources has reached an extremely high level of quality, but in most cases, we see only philologists at work, with the consequent publication of excellent digital editions, where the historian often works traditionally. In the other two major areas – GIS and the web – the situation is much better, but only if we look at the relative weight of these large areas compared to all the methods that emerged from the world of the digital humanities, and not because of their relevance in the historical studies.

This 'distance' highlights a serious problem. A long time has passed since the first pioneering experiments in the field of the Digital History and skepticism has been growing ever since. It is no longer possible to carry out historical research without knowing the methods offered by the digital technologies in order to process information and – as Serge Noiret says – "we can hardly imagine separating historical research from the tools, practices and programs necessary to carry it out". This, in fact, "is no longer a viable road" (2019, 12).[26]

25 The essay about the software MACHIATO is meaningful for the understanding of Machiavelli's diplomatic correspondence described nowadays still as a "potential" with all the "dangers" of a "militant" initiative (Manchio 2019, 207-26).

26 A very old question already posed by Manfred Thaller (1985, 871-90) and Robert Rowland (1991, 693-720).

e-ISSN 2724-3923

My analysis simply confirms a well-known problem, which in Italy has not been tackled seriously so far: the education of the historians in the use of digital tools required for his/her profession. There is an almost total absence of Digital Humanities courses and programmes in the bachelor and master's courses of history in Italy. This has in fact put Digital History in an "out of the box" position and placed historical researchers in friction with the digitization of our whole society and everyday life.

The spread of the digital in every filed of our lives, combined with the lack of a suitable education, has pushed the historians to spontaneously – and therefore haphazardly – approach only the digital methods and tools perceived as immediately useful. As regards the varied and vast world of the web, this path has led to a partial digitization of the tradition, both for research and dissemination; with the exception, for the reasons already mentioned, of the digital public historians.

As far as GIS is concerned, the question is more complicated. The use of GIS in digital public history implies – if the tool is used only for the visualisation of historical data – the acquisition of specific skills. GIS requires a huge construction work of the data model, a wise design of the platform allowing for a dynamic visualisation, transparency on the methodological choices, and a detailed documentation to restrain the user's disorientation in front of the search engine.[27] It is clear that GIS is a really important tool for the historian as well as a demanding one: hence also the advantages of the application of GIS in digital history are confined to a small niche of users.

If Italian universities continue to churn out historians unaware of the main methodological questions posed by the digital humanities and unable to master their main tools, the "out of the box" position of digital history is destined to consolidate, no matter how much interesting, useful, and methodologically valid work is carried out by these self-formed digital historians. Back in 2015 Serge Noiret said: "today the 'digital' part of the historian's profession has become essential even when one does not think of practicing a new discipline such as 'digital history' within digital humanities, but of continuing traditional practices updating them" (2015, 267). However, this update nev-

27 From this point of view an excellent example is Slave Voyages (`https://www.slavevoyages.org`) created by the Emory Center for Digital Scholarship, the University of California (Irvine and Santa Cruz), and the Harvard University Hutchins Center: the result of a three-year (2015-2018) work carried out by an interdisciplinary team of cartographers, computer scientists, historians, librarians and web designers through the collaboration of scholars from various European, African and American universities, Slave Voyages does not 'only' allow to explore and display in a unique dynamic multi-source dataset on the transatlantic travel of slaves, but offers a rich set of explanations of sources, data model and historical essays that help the research and the interpretation of the phenomenon.

e-ISSN 2724-3923

er took place,[28] and this took history away from academic historians, opened their profession to an extremely wide and varied range of people and institutions, and left the historians outside the process of developing new methods of analysis. The trend seems clear, and while not necessarily negative or worrying, it is a position that has serious consequences on the relationship among the university, society, and the job market.[29] So far, nothing has been able to change this standoff: not the pioneering works – now good practices after forty years of experimentation –, nor the articles that highlighted potential and problems, nor the research that explained the innovations brought by the digital tools. The only relevant countertrend signal could come, in my opinion, from the Public History movement and its inevitable digital component. Although there is ongoing resistance of the Italian academy to incorporate the main tools and methods of the Digital History, new scenarios could be opened in Italy for the digital historian if Public History adds these tools to the historian's educational path.

6 History (with the Digital) and the Problem of the Statements

My previous statement could run counter to those who make a distinction between Digital History and Digital Public History as well as between Digital History and 'history with digital tools'. Serge Noiret says that it is important to define the respective areas, in order to better highlight the characteristics of the Digital History in the wide galaxy of the Digital Humanities: "digital history is not history with digital and it is no longer time of generalist fields and universal humanistic practices with digital", and "the digital history that uses and dominates technologies always refers to specific cognitive practices of historians and of the historian's job". Consequently, Noiret differentiates between "research, teaching, communication of the outcome [of historical research] today necessarily linked to the digital" (i.e 'history with digital tools') and "digital history" strictly speaking (2019, 13). The latter is defined by Deborah Paci as "a research

28 Many international authors denounce this lack of evolution everywhere. For example, Toni Weller (2013) talks about the soft impact of the digital revolution on the pre-existing practices of historians and in continuity with their professional traditions. Technology has not led to a new discipline from which to move in order to solve epistemological problems that, without digital tools, could not even be thought of.

29 A short personal note: in 20 years of teaching in Pisa in the degree course in Digital Humanities, several of my students – with an 'insufficient' historical education but a good one in Digital Humanities – found good jobs in archives, museums, libraries and in research projects with relevant historical-cultural aim, while their colleagues from the degree courses in History or Cultural Heritage were struggling to find a job corresponding to their CV.

area that uses, in the scholarly field of historical disciplines, methodologies, computational tools and computer techniques aimed at automatic or semi-automatic data processing, which are displayed and given back to the scholar through quantitative analysis" (2019, 19).

Personally, I find the definition of Deborah Paci correct, but partial, because the wide and articulated range of complementary activities that leads the historian to be 'digital' does not necessarily involve quantitative analysis, but all application of digital technologies to historical research that are methodologically sound. To avoid misunderstandings, I do not think that consulting the MGH online, publishing one's essay on `academia.edu`, creating a bibliography on Zotero or broadcasting a history conference on streaming channels makes a historian digital, but maybe all these things together in a unified and well set project would. However, the profession of historian includes a very wide range of activities, in which digital tools and methods have a relevant place even if they do not involve computational activities.

There are numerous fields in which digital technology enters the work of the historian without necessarily involving an "automatic or semi-automatic processing of data" or, at least, in which such treatment constitutes only a part of the process. The historian's job is never confined to pure research, isolated from the environment, but lives on the deep interconnections that it has with the society which it communicates with. In the Public History Tree designed in 2019 by Thomas Cauvin these interconnections are expressed in an extremely effective way:

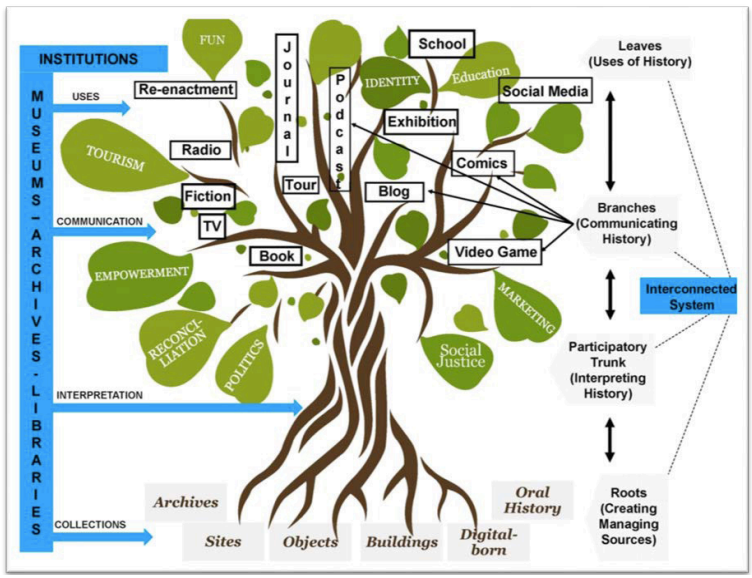

Figure 3 Public History Tree designed in 2019 by Thomas Cauvin

> The tree is divided into four parts: the roots, the trunk, the branches and the leaves. Although they are different, these parts belong to a single system; one cannot exist without the other. While history has traditionally been defined as the rigorous and critical interpretation of primary sources (the trunk), public history is somewhat broader and includes four parts. The roots represent the creation and conservation of the sources; the trunk corresponds to the analysis and interpretation of the sources; the branches represent the diffusion of these interpretations; and the leaves are the multiple public uses of such interpretations. The more connected the parts, the richer and more coherent public history becomes. Furthermore, the structure is not linear; the uses (leaves) often have an impact on what we consider important to collect and store (roots). The public tree should not be seen as a pure linear process, but rather as an interconnected system. (Cauvin 2020, 20-1)

Cauvin's tree schematically represents the Public History that is – in reality – a subset of History as a discipline. But I do not think that the differences – between Public History and History – are so macroscopic. Many believe that traditional history is localised in what we could define – in Cauvin's tree – as the roots and the trunk, i.e. in the exegesis. This is characterised by the comparison of the sources, new interpretations, and communication of the outcomes only or mainly to the scholarly community. Even if for someone the tree of History may look more like a cypress – with reduced and codified interactions with the public – than like an oak or a willow, the interconnection between all parts remains strong and tight, because the historian must always use digital tools with a strong awareness, dominating and understanding their mediums. A digital historian working on the social media may certainly have to use techniques relating to the world of big data, but first of all, he/she must understand and dominate the medium itself, he/she should use social media to share information, interact with its audiences, collect the outcomes of this interaction and then make them flow back into new research. Making a digital edition of a historical source means not only knowing the 'text encoding' tools and methods, but also developing an appropriate encoding that allows the recovery and the comparison of historical information in order to allow for specific visualisations and analyses.[30] The digital historian must also decide on 'how' the user can access information and what information to convey, because the

30 For example, the variable spelling of common or proper names in medieval sources is a piece of information that is interesting in itself but that also requires a standardisation in order to search both variations (linguistic) and the meaning (history). In this regard, see also Thaller 2017.

e-ISSN 2724-3923

impact of that decision will necessarily influence the individual's historical interpretation of the source itself and that of the various communities (scholarly, local, specialised etc.).

Could I make 'digital history' without being aware of it? Could I build the website of my research not knowing the problems related to the concepts of original, copy, authenticity, counterfeiting, distribution, conservation, forgery that the digital inevitably places on the treatment of sources? Could I open the edition of a text to external comments and to a shared interpretation without being aware of the issues related to the shared authority on the web? If I do have all this knowledge, am I, or am I not a digital historian?

I know I will ruinously fall into a tautology, but I would simply modify the definition of Deborah Paci in this way: 'digital history is a research area that employs, in the scholarly field of historical disciplines, methodologies, tools and IT techniques aimed at effectively answering historical questions'. Again Serge Noiret wonders "whether or not to continue referring to digital history today in the field – really generalist umbrella – of digital humanities" and then concludes that it is "better to translate digital humanities into individual disciplines, not to confuse tools, methods and questions". The question is, in my opinion, ill posed. Certainly "as historians we need to create content" and have a clear "originality of our methods, tasks and final objectives in the digital field", but we cannot do it using "digital tools different from those used by other humanists, who above all promote the exegesis, analysis and codification of the text" (Noiret 2019, 14).

There are no 'different digital tools'. The digital historian – as we have seen – can better recognise which tools and methods belong where and which answer other scholarly questions. The historian should also use the tools of the Disciplines of the Text. Text analysis works really well and should still be used to answer to historical questions.

The return to the individual traditional disciplines would, in my humble opinion, be the real gravestone on the revolutionary flow of the digital turn for the humanistic field, which historians have often deserted. It is on the originality and specificity of the historical question – not on the tools – that the digital historian should insist on. Within the world of the digital humanities, above all, there should be a strong collaborative and interdisciplinary dimension with other digital humanists.

Several projects in the field of digital humanities are potentially relevant for the historians. These projects do not usually have historians in the team and are often incomplete and unsatisfactory. They offer large amounts of data and of resources that confuse the common reader and sometimes manage to answer only the questions raised by the creator of the resources.

Marco Tangheroni used to say that historical sources are like 'well-behaved girls', they only answer if you ask them. The answer – per-

haps – lies precisely in thinking about the historical questions and in investigating without fear of interdisciplinary mixes. We must use all the available tools, update them, modify them, and adapt them to the questions themselves. This operation requires a radically new way of working as an interdisciplinary team. Academic historians are very reluctant to adopt this approach, but for the public (digital) historian, maybe, it is obligatory.

Bibliography

Allegrezza, S. (ed.) (2019). "Didattica e ricerca al tempo delle Digital Humanities". *AIUCD2019 - Book of Abstracts*. http://amsacta.unibo.it/6361/.

Burgio, E.; Simion, S. (a cura di) (2015). *Giovanni Battista Ramusio. Dei viaggi di Messer Marco Polo. Edizione critica digitale progettata e coordinata da Eugenio Burgio, Marina Buzzoni, Antonella Ghersetti*. Venezia: Edizioni Ca' Foscari. http://virgo.unive.it/ecf-workflow/books/Ramusio/main/index.html.

Cauvin, T. (2018). "The Rise of Public History: An International Perspective". *Historia Crítica*, 68, 3-26. https://doi.org/10.7440/histcrit68.2018.01.

Cauvin, T. (2020). "Campo nuevo, prácticas viejas: promesas y desafíos de la historia pública = New Field, Old Practices: Promises and Challenges of Public History". *Hispania Nova. Primera Revista de Historia Contemporánea online en castellano. Segunda Época*, 1 extra (7/5/2020), 7-51. https://doi.org/10.20318/hn.2020.5365.

DanteSources. Per una enciclopedia dantesca digitale (2016). https://dantesources.dantenetwork.it/.

Edizione digitale dell'Epistolario di Alcide De Gasperi. https://www.epistolariodegasperi.it/#/.

Gil, T. (2015). "Storici e informatica: l'uso dei database (1968-2013)". *Memoria e ricerca*, 50, 161-78. https://doi.org/10.3280/mer2015-050010.

Gil, T. (2019). "L'utilizzo dei database da parte degli storici: storiografia e dibattito attuale". Allegrezza, S. (a cura di), *AIUCD2019 - Book of Abstracts*. Udine, 2019, 177-81.

Italia, Paola (a cura di). *Idilli di Giacomo Leopardi*. http://leopardi.ecdosys.org/it/Home/.

Manchio C. (2019). "Per un'analisi 2.0 della corrispondenza machiavelliana". Paci 2019, 207-26.

Marras et al. (a cura di) (2020). *La svolta inevitabile: sfide e prospettive per l'Informatica Umanistica*. Milano: Università Cattolica del Sacro Cuore, 2020. http://amsacta.unibo.it/6316/.

Maxwell, A. (2019). "Le ricerche storiche sugli archivi digitali. Alcune note di ricerca". Paci 2019, 161-74.

Noiret, S. (2009). "'Public History' e 'storia pubblica' nella rete". *Ricerche storiche*, XXXIX/2-3 (2009), 275-327.

Noiret, S. (2011). "'Public History': una disciplina fantasma?". *Memoria e Ricerca*, 37, 10-35.

Noiret, S. (2015). "Storia contemporanea digitale". Minuti, R. (a cura di), *Il web e gli studi storici. Guida critica all'uso della rete*. Roma: Carocci, 267-300.

Noiret, S. (2019). "Prefazione. Homo digitalis". Paci 2019, 9-18.

Paci, D. (2019). *La storia in digitale. Teorie e metodologie*. Milano: Edizioni Unicopli.

PoLet500. http://www.polet500.it.

Rowland, R. (1991). "L'informatica e il mestiere dello storico". *Quaderni Storici*, XXVI, 78, 693-720.

Santarelli, D. (a cura di) (2019). *Invito alla storia: terza conferenza italiana di Public History* (Santa Maria Capua a Vetere e Caserta, 24-28 giugno 2018, AIPH, 2020). https://aiph.hypotheses.org/9076.

Salvatori, E.; Privitera, C. (a cura di) (2018). *Metti la storia al lavoro: seconda conferenza italiana di Public History* (Pisa, 11-15 giugno 2018). https://aiph.hypotheses.org/7389.

Salvatori, E. (2018). "Per un'analisi delle pratiche di Public History". Salvatori, Privitera 2018, abstract 47.

Spampinato, D. (2018). "Patrimoni culturali nell'era digitale". *AIUCD2018 - Book of Abstracts*. http://amsacta.unibo.it/5997/.

Storey, H.W.; Walsh, J.A.; Magni, I. (eds). *An Edition of Petrarch's Songbook Rerum vulgarium fragmenta*. http://dcl.slis.indiana.edu/petrarchive/.

Tammaro, A.M. (2005). "Che cos'è una biblioteca digitale?". *DigItalia*, 1. http://digitalia.sbn.it/article/view/325/215.

Thaller, M. (1985). "Possiamo permetterci di usare il computer? Possiamo permetterci di non usarlo?". *Quaderni Storici*, XX, 60, 871-90.

Thaller, M. (1993). "Historical Information Science: Is There Such a Thing? New Comments on an Old Idea". *Historical Social Research Supplement*, 29, 60-286.

Thaller, M. (2017). "Between the Chairs. An Interdisciplinary Career". *Historical Social Research Supplement*, 29, 7-109. http://doi.org/10.12759/hsr.suppl.29.2017.7–109.

Tomasi, F. (2013). *Vespasiano Da Bisticci's Letters. A Semantic Digital Edition*. http://doi.org/10.6092/unibo/vespasianodabisticciletters.

Vega, Lope de (2015). *La dama boba: edición crítica y archivo digital*. Bajo la dirección de Marco Presotto y con la colaboración de Sònia Boadas, Eugenio Maggi y Aurèlia Pessarrodona. PROLOPE, Barcelona; Alma Mater Studiorum, Università di Bologna, CRR-MM, Bologna.

Vitali, G.P. *Last Letters from the World Wars: Forming Italian Language, Identity and Memory in Texts of Conflict*. http://www.ultimelettere.it/LastLetters/people/.

Weller, T. (2013). *History in the Digital Age*. Routledge.

e-ISSN 2724-3923

magazén
Vol. 1 – Num. 2 – December 2020

Digital Heritage Consumption: The Case of the Metropolitan Museum of Art

Trilce Navarrete
Erasmus University Rotterdam, Nederland

Elena Villaespesa
Pratt Institute, School of Information, New York, USA

Abstract Cultural consumption is increasingly moving into a digital realm where art and non-art spaces blur in an all-inclusive image-rich environment online. While cultural consumption studies remain limited to a defined cultural environment (e.g. the museum website), we will analyse the patterns of consumption of 119 paintings from the Metropolitan Museum of Art collection in an all-inclusive online environment, Wikipedia. We will find paintings in art as well as non-art related articles and compare visibility to the institutional physical and online exhibition, a purely art environment. We will find a greater share of digital cultural consumption takes place in non-art related articles, inferring accidental consumption, while fact-checking and the presence of articles about obscure paintings satisfy a utilitarian information use. We will argue that digital cultural consumption can expand the user base when positioned outside of the expected art context, enabling new forms of hedonic and utilitarian consumption. Our results suggest that the adoption of the online encyclopaedia by superstar museums reflects a new conceptualisation of authentic taste that includes digital consumption, highlighting the collection's information value.

Keywords Cultural consumption. Digital paintings. Museums. Wikipedia.

Edizioni Ca'Foscari

Peer review

Submitted	2020-10-09
Accepted	2020-11-12
Published	2020-12-22

Open access

Citation Navarrete, T.; Villaespesa, E. (2020). "Digital Heritage Consumption: The Case of the Metropolitan Museum of Art". *magazén*, 1(2), 223-248.

DOI 10.30687/mag/2724-3923/2020/02/004

1 Introduction

Museums are organisations with limited budgets and growing collections serving an increasingly digital consumer who expects collections to be remotely available. The International Telecommunications Union (ITU) released the 2017 global Information and Communication Technology (ICT) figures showing more than half of households have access to the internet, with mobile broadband dropping in price and growing in subscriptions to reach 4.3 billion users, with 70% of the young (15 to 24 years old) being online (ITU 2017). This increasingly digital context raises the need to identify the characteristics of the online consumer to evaluate the relevance of a cultural digital presence, which looks grim as audience diversity appears to only decrease online (Mihelj, Leguina, Downey 2019). It must be said that, while museums are investing on their websites, the use of social media, and a number of mobile applications to enhance visitor experience (Zins 2107), access to collections online continues to lag behind. Museums reported a third of their collections have been digitized yet less than 10% are published online (Nauta, Van den Heuvel, Teunisse 2017; Axiell 2017). This may explain the limited growth, and even decline, in the use of museum websites for accessing collections (Mihelj, Leguina, Downey 2019). Consumption of museum's services continues to be foremost a physical experience, to a specific small group of society. Museum visitors follow a specific socio-economic profile (e.g. Falk, Katz-Guerro 2015) and enjoy a relative ease to physically access institutions (Brook 2016; Evans 2016). Alternative channels to disseminate collections have been considered in an effort to broaden the visitor profile, increasingly including third party platforms such as Artstor, YourPaintings, Google Arts and Culture, and Wikipedia. Katz-Guerro (2004) calls for an inclusive approach to understanding cultural consumption, considering "a variety of alternative forms and styles that reflect different theoretical approaches and different facets of the life-world" (13). She identifies Internet use of cultural content as an understudied area of research. Our study responds to fill this gap.

In this paper, we will analyse the consumption of paintings when museums disseminate collections via Wikipedia, one of the 10 most visited websites globally for over a decade (Navarrete, Borowiecki 2016; Spoerri 2007b). We will focus on the paintings collection of the Metropolitan Museum of Art in New York (The Met), our case study, and analyse their use as illustrations in Wikipedia. We will manually code the topic of the articles where paintings are found to rank the frequency of use and views, to finally compare the visibility of the paintings within the institutional website and exhibition halls. Results will show a pattern of utilitarian consumption that respond to a clear information need, where paintings are used as visual documentation and authentic information sources, not only as art works.

Whereas onsite visitors are generally considered hedonic consumers, the online use of collections reflect a utilitarian approach to museum collections. The main contribution of this paper lies in identifying the variety in non-art contexts where consumption of collections takes place, illustrating Wikipedia articles. By the mere size of online consumers, it can be assumed that online visitors include museum non-visitors, broadening the consumer base. This suggests that museums that make their paintings available as illustrations of the online encyclopaedia exponentially increase the view to their collections by enabling consumption of paintings as art and as information, hence doubling potential utility.

The remaining of the paper is organised as follows. Section two contains the review of literature on cultural consumption, noting the digital variant and the studies on availability of cultural content, as well as literature available on Wikipedia consumers. Section three presents the methodology and data, while section four discusses the findings. We close with conclusions and policy implications.

2 Cultural Consumption: Who Visits Museums and Why (Not)?

Museums are institutions holding the most representative objects from our past. Objects are carefully selected, preserved, researched and exhibited in special buildings designed to protect and showcase collections. Collections are cultural goods, which are complex experience goods that may present addictive behaviour leading to greater utility (Frey, Meier 2006). That is, the more one visits museums the more enjoyable they become. Most studies take cultural consumption as the dependent variable to identify the relevant socio-economic characteristics that will lead to a museum visit. A European study by Falk and Katz-Guerro (2015) found that per capita household income, education, labour market status, and country of birth are important determinants to positively influence museum visits across countries. In an earlier study, Ganzeboom (1982) identified the information-processing capacity of individuals to be a key determinant in the choice of culture consumed. He argued that a certain complexity of cultural products leads to a rewarding experience for the consumer but to a certain point, as greater complexity does require greater human capital (skills and knowledge). That is, consumer utility is directly related to the individual's information-processing capacities. DiMaggio (1987) argued that cultural consumption is actually a form of currency to establish one's identity in any social encounter. In this way, individuals exchange information about their cultural taste and form social networks accordingly. Generally, the greater the cultural vocabulary, or diversity in taste, the higher the socioeconomic status.

Television serves to provide "a stock of common symbols for nearly everyone" which can then be exchanged in social encounters (443), represents a limited source of cultural content and has been associated with individuals having narrow social networks (e.g. low-income, blue-collar, unemployed, aged).

However, taste cannot only be accounted for as a dynamic for class differentiation. Instead, Meyer (2000) proposes institutions to be central players in contextualising the collective definition of taste, specifically along the rhetoric of refinement and the rhetoric of authenticity. Both concepts are inherited from a European aristocratic context, which serve to define and legitimise contemporary standards of taste. According to Meyer (2000), refinement defined the practice and mentality of the exclusive aristocracy and served to differentiate it from the lower classes, while authenticity was proposed to challenge a notion of taste that appeared to be based on inaccessibility and cost: "wherever taste is expensive, it is false" (Rousseau quoted in Meyer 2000, 42). Authenticity became hence the definition of true taste. In contemporary society, both notions of taste are complementary, albeit at times conflicting. Museums can serve to exemplify the role of the institution that defines a changing context of taste, for it is museums that define object value and meaning (Cameron 2007). In the case of photography as art form, artists were involved in the development of photography and adopted it as an art form early on, but it was the museum institution that established photography as authentic art form (Walsh 2007; Fyfe 2004). It is to be expected that museums follow a similar process to define digital reproductions of their collections, along digital art, as authentic. Museums hence form our notions of taste and shape our cultural consumption pattern (Gans 1974 in Katz-Guerro 2004; Meyer 2000; Beckert, Aspers 2011).

Cultural consumption is first and foremost limited by accessibility of available content. It has been estimated that less than 10% of museum objects are available for view (Frey 2000) which is understandable when considering the size of the collections in comparison to the available wall space to exhibit them. However, availability of content through digital exhibition is surprisingly similar. European museums reported having digitised about a third of their collections, of which less than a third is available on the Internet (Nauta, Van den Heuvel, Teunisse 2017). That means, less than 10% of their collections is available for digital consumption. Physical accessibility to the museum building was found by Brook (2016) as key determinant for consumption in her study on museum visitors in London. When physical accessibility to cultural services is improved, consumption increases, both of the expected higher socio-economic groups as well as the rest of the population (Brook 2016). This result rises the expectation that digital availability would increase consumption of all socio-economic groups. This does not appear to be the case, as consumers on-

line tend to reproduce the current museum visitor profile, and even enlarge the participation divide (Evrard, Krebs 2018; Mihelj, Leguina, Downey 2019, Finnis, Sebastian, Clemens 2011).

Besides accessibility of cultural services and availability of museum collections, both physical and digital, what are the barriers for greater cultural participation? A study on barriers to cultural participation in the United States shows that lack of time (55%) and lacking somebody to go with (22%) are important reasons for not visiting an art exhibit (NEA 2015). Europeans reported lack of time (37%), lack of interest (31%), and lack of information (25%), to be the important reasons for not visiting museums (EU 2017a). It can be argued that these are related, as lack of interest results from a lack of information since utility of cultural consumption is known to be linked to previous consumption, and we all know we make time for things that interests us. Jarnes (2015) proposes an alternative framework to understand new forms of emerging cultural capital, as cultural content gains new distribution forms and individuals consume across established social structures. He compares cultural and material consumption in Norway and identifies four consumption profiles: intellectual, luxurious, educational, and practical (see Table 1). Curiously, they all attend the symphonic orchestra yet the motive to visit, as well as the classification and evaluation of the performance, follows a different reasoning.

Table 1 Consumption profiles

Profile Characteristics	Intellectual	Luxurious	Educational	Practical
Education	MA and PhD, often in humanities.	BA not in humanities.	BA or lower.	Below BA.
Capital	High volume, mostly cultural. Work in public sector or non-profit.	High volume, mostly economic. Work in private sector.	Heterogenous, average wage. Work across public and private sectors.	Homogenous, low overall capital. Working class in private sector.
Cultural consumption	Art for art sake. Prefer new, experimental and challenging art.	Utilitarian, must relax and entertain. Prefer high end art.	Moralistic, must increase knowledge, present the truth.	Meaningful, must be comprehensible. Prefer documentaries and high tech.
Material consumption	Moderate and ascetic to abstinence.	Luxury for luxury sake, sensitive to brands.	Materialism is seen as squandering. Excessive consumption is immoral.	Pragmatic, inconspicuous, distrust posh.

Source: Based on Jarnes 2015.

e-ISSN 2724-3923

These four singular profiles provide an alternative view into the potential interest and reasoning behind cultural consumption. Of particular interest are the educational and practical profiles proposed by Jarnes (2015), when analysing the visibility of paintings in an online encyclopaedia, as they are both interested in culture when it can educate, increase knowledge, and inform about a topic, with a touch of high tech but may not necessarily visit a museum. Seeking culture online is not always easy if consumers are required to perform a search because they may lack the cultural knowledge required to do so. Finding the Anatomy Lesson of Dr. Tulp by Rembrandt housed at the Mauritshuis requires knowing the title, the painter, or the museum where the painting is located. In a study of 50 websites from the cultural domain, Stiller (2012) highlighted the fact that consumers can benefit from serendipity and exploration to encounter cultural content online, which is often lacking in culture-related websites. Instead, general-purpose websites, such as the online encyclopaedia Wikipedia, are ideal points to encounter with cultural content through serendipity, exploration, or mere accident. In this way, readers of the article "Autopsy" would accidentally encounter the painting by Rembrandt.

In the meta-analysis by Wu and Lu (2013), a distinction was made between hedonic consumption (for pleasure or interest causing satisfaction), utilitarian consumption (goal-driven and dependent on environment), and dual-purposed (a combination of both). Many websites and services online were categorised as dual-purposed, as a specific information need was complemented by a certain degree of satisfaction. Intrinsic motivators were found more critical for hedonic and for dual-purposed use, highlighting the significance of pleasure and enjoyment in systems to access information. Considering the case of the Wikipedia website, the study by Navarrete and Borowiecki (2016) found greater views of articles containing a larger number of images from a diversity of sources. The authors link this observation to a signal of quality, suggesting utilitarian digital cultural consumption, while we argue that the presence of art images in encyclopaedic articles may play a different role depending on the context.

2.1 Consumers of Wikipedia

Wikipedia is the online encyclopaedia that was launched in 2001 as main project of the Wikimedia Foundation, currently including a number of other complementary projects. The website gives free access to crowdsourced content of articles in over 250 languages, positioning it as one of the top 10 websites worldwide according to Alexa online ranking. The English Wikipedia has 5.7 million articles and receives 4.5 million views per hour worldwide (`http://stats.wiki-`

media.org/EN). There are numerous studies on the motivation of unpaid contributors, the management of content, the patterns of reader consumers, as well as applications of content towards information retrieval and language processing systems. Science direct currently identifies over 12,000 articles about Wikipedia while Wikipedia itself lists over 70 books and book chapters, 4,000 conference papers, 1,500 articles, 52 doctoral thesis, and several other scientific output (http://wikipapers.referata.com/wiki/List_of_publications).

Wikipedia has been studied as an example of a knowledge commons (Hess, Ostrom 2006), with a substantial body of work around the organisation and motivation of editors (and bots) as well as the evaluation of the content (for an overview of literature see Julien 2012). Evaluating consumer (reader) satisfaction generally relies on article frequency of views, assisted by a number of available tools developed by the Wikimedia community (see https://stats.wikimedia.org/). One approach to the analysis of reader preference can be found in Spoerri (2007a, 2007b) who looked at the 100 most visited pages and their positioning over time. He developed a primary and secondary category system to organise the articles to examine, if indeed, the expected geography, history, and science encyclopaedia categories were the most visited pages. His results show nearly half of all views in the five-month period to represent the category of entertainment, and an unexpected 10% views to the category of sexuality, the latter constituting thirty percent of the most visited pages in the sample. Spoerri (2007b) further established the relation to the most popular queries submitted in search engines, where pages from Wikipedia were ranked in the top three results in Google, increasing popularity to Wikipedia. This trend in fact changed after 2012 when Google introduced the knowledge graph to include the basic data from Wikipedia at the top right box of the results page (fed by DBpedia, extracted structured data in Wikipedia), reducing traffic to Wikipedia articles (https://en.wikipedia.org/wiki/Knowledge_Graph).

An application of the same categorisation can be found in Reinoso et al. (2012), who excluded Religion, Holidays, and Drugs in their analysis of the most visited Wikipedia pages in the four largest languages, namely German, English, Spanish and French Wikipedia. Their results of the English Wikipedia support the findings by Spoerri (2007a, 2007b), where Entertainment received nearly half of the views and Sexuality received the expected ten percent of views. Because Wikipedia lacks a systematic category system to organise articles, we have adopted the categorisation system devised by Spoerri (2007b) for the analysis of extracted data, being the available categorisation of the online encyclopaedia articles found in related literature. Other authors who examined popularity are Ratkiewics et al. (2010a, 2010b). Their approach was to examine the popularity of Wikipedia articles in relation to queries in the Chilean web. Their

e-ISSN 2724-3923

results point to a characterisation of a heavy-tail burst behaviour influenced by endogenous events, such as an Oscar nomination for an actor's page or election day for a candidate's page. Similarly, Mestyan, Yasseri and Kertesz (2013) used Wikipedia popularity to predict movie box office success.

Limited empirical evidence exists on the use of and access to museum collections in Wikipedia. One example includes the analysis of the visibility of collections from the National Ethnographic Tropenmuseum in Amsterdam (Navarrete, Borowiecki 2016). Results showed that online publication of collections resulted in a substantial increase in object visibility compared to onsite exhibition of objects, as well as difference in consumer preference for type object, favouring 3D objects onsite and 2D objects online. A more recent study analysed 8,000 paintings used in 10,000 articles in the English Wikipedia, following a similar categorisation of articles, where 33% of articles were art-related though receiving 12% of views, while 67% of articles containing a painting were non-art related and received 88% of views (Villaespesa, Navarrete 2019). The authors argued the legal framework of open data to be responsible for the use of and access to collections.

A survey to Swiss museums revealed awareness of the potential increase in access to collections when disseminating open images via the online encyclopaedia, to stimulate education and research. However, institutions lacked resources to embark in the process to publish collections as open data and were reluctant to allow the inevitable commercial use of collections (Estermann 2013). In contrast, costs to make collections available in Wikipedia were considered negligible by the Dutch Tropenmuseum, as all the images and metadata were already available and transfer was simple (Estermann 2013). Successful collaborations between museums and Wikipedia are many, including the Derby Museum in the United Kingdom that included QR codes with a link to a Wikipedia article where visitors could learn more about the topic.[1]

2.2 Available Data on Digital Cultural Consumption

Even though digital technology has been identified as a threat to further engender equality of access, as digital literacy and access to technologies form a new determinant for Internet-related consumption (Katz-Guerro 2004; Ateca-Amestoy, Castiglione 2016), little research has been done on the actual emerging consumption pattern

1 A list of 22 selected successful case studies can be found at `https://outreach.wikimedia.org/wiki/GLAM/Case_studies/Archived`.

e-ISSN 2724-3923

of paintings online. There are indications that digital consumption is complementary to onsite visits to museums (Ateca-Amestoy, Castiglione 2014; Evrard, Krebs 2018; Mihelj, Leguina, Downey 2019), as well as for art sales (Hiscox 2017). There are a number of case studies on the development of new art and culture services for mobile use (EU 2017b). Data available on cultural consumption (EU 2013) generally focuses on the so-called high art forms, where roughly a third of European adults visit museum (37%) and half visit heritage sites (52%). Translating the question online proves to be challenging. Digital consumption of music may be easier to understand when comparing visits to live concerts (35%) or listening to music on radio or television (72%) to listening to radio online (42%), downloading or streaming music (31%), or buying CDs online (27%). Respondents of visiting museums instead report "visiting museum, library or other specialised websites to improve your knowledge" (24%), or a more general "searching for information on cultural products or events" (44%). The hedonic side of consumption does not appear to be represented for museum collections in the online survey variant even though museums are engaging in the dissemination of their collections via Facebook, Instagram or SnapChat in order to extend the dialogue beyond a museum's physical location (Weilenmann et al. 2013). Museum content can be consumed separated from a visit to the institution just as a song can be heard independently of the programming of a concert hall.

Understanding the social consumption patterns of views to museum objects remotely via the Internet has received little attention, perhaps due to three central methodological challenges: (1) categorising remote digital access of museum collections has yet to be considered a form of cultural consumption; if this was to happen, (2) agreement on a harmonised method to measure consumption (e.g. views, clicks) is required; in order to (3) build datasets to allow an analysis of consumer behaviour, regular updates and revisions of the scope of the medium (e.g. images, XR) and the distribution channels (e.g. Wikipedia) should be ensured. In this paper, we argue that cultural consumption does not have to take place explicitly within a cultural context, such as a museum or a museum website, but that it can also take place within an information context, as accidental cultural consumption (e.g. in Wikipedia), both as hedonic and utilitarian forms of consumption. We further propose viewing paintings used to illustrate Wikipedia articles as an alternative consumption of museum content. Such consumption is not comparable to a physical or virtual visit to a museum but, in fact, points to the information value of collections that serve as visual documents to the encyclopaedic articles. Consumption of museum collections would in turn gain an entire new share of the market to increase the user base, beyond the traditional museum visitor.

3 The Case of the Metropolitan Museum of Art Paintings on Wikipedia

The Metropolitan Museum of Art (The Met) was founded in 1870 and is located in New York City, where it received 7 million visitors in 2017, of which 30% were locals and 37% were internationals. The museum has an online collection of over 400 thousand objects, of which over 13 thousand are paintings (The Met 2017).[2] The general website of the museum receives 32 million yearly online visits while the collection online receives 7.2 million yearly visits. These 600 thousand monthly visits can be categorised as follows: professional researchers, students, personal interest information seekers, casual browsers, information seekers and visit planners (Villaespesa 2017). In 2017, the museum launched an open data policy and, at the time of writing, 216,636 objects (of which 8,691 paintings) have been made available online with high-resolution images under Creative Commons Zero (CC0). The museum collaborated with the Wikimedia Foundation and uploaded all 364 thousand images as part of the Open Access Initiative (The Met 2017). A review of the results one year after the launch of the Open Access strategy showed an increase in image downloads from the museum's website, a higher usage of object images on Wikipedia, and a 385% increase in page views of articles that included an image from the collection (Tallon 2018).

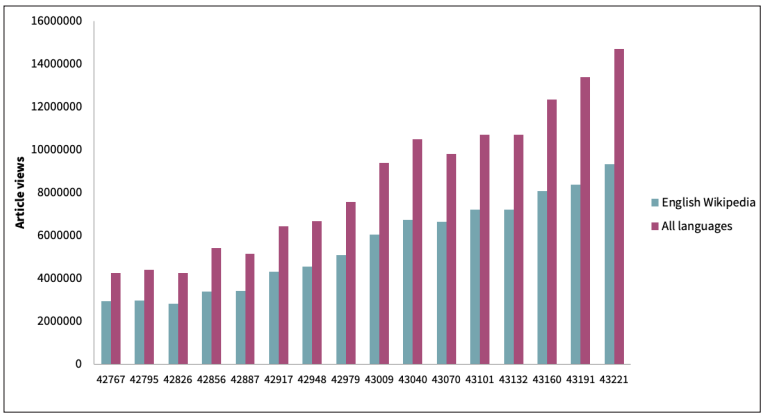

Figure 1 Number of Wikipedia articles views that include an image from The Met

2 The collection can be access online at `https://www.metmuseum.org/art/collection/`.

This study analyses a section of the collection, namely paintings, that have been used to illustrate articles in the English version of Wikipedia. The dataset corresponds to the month of May 2018 and has been gathered using the BaGLAMa2 tool,[3] which shows all the image metrics for the category "Images from the Metropolitan Museum of Art". These categories were made to facilitate the tracking of content use, often in collaboration with the museum institutions, but are not compulsory and hence not exhaustive. Our dataset comprises 119 paintings used in 169 Wikipedia articles receiving a total of 2,589,378 views in May 2018.

Table 2 The collection of the Metropolitan Museum of Art

	In museum	Open Access	In Wikimedia Commons	As illustration of articles	Number of articles
Total collection	461,591	216,636	364,359	NA	1262
Total paintings	13,269	8,691	NA	119	169

NA = Not available.

From the 364,359 object images the Metropolitan Museum of Art has shared in Wikimedia Commons for use in articles, we identified 119 painting images were actually used. We selected those articles that included a painting and that received more than 500 monthly views for further analysis, representing a total of 169. Those articles were manually coded using the categories found in the literature (Spoerri 2007a, 2007b). We added a category of Wikipedia that comprises all the pages created by the encyclopedia that are not necessarily articles (e.g. Home page, Featured article, User profile). These were removed from the analysis to focus on the articles and to reduce potential temporary variations in the average traffic. Results show a significant share of the paintings is used in art related encyclopedia articles, representing 57% of the sample, followed by geography (17%) and history (16%). As seen in Table 3, the views received by article class rank a bit different. It is not surprising to see the art-related articles receive the largest share of views (33%). What we found surprising is that history nearly ranked as high with 32% share of total views.

3 The BaGLAMa2 tool was developed by Magnus Manske, to track the number of human views to Wikipedia articles from a specific commons category. The category representing the Metropolitan Museum of Art collection comprises 18 months (`https://tools.wmflabs.org/glamtools/baglama2/index.html#gid=290&month=201807`).

Table 3 Number of articles and number of views per category (N=169)

Primary category	Articles	%	Article views	%
Arts	96	57%	855,638	33%
Geography	29	17%	440,669	17%
History	27	16%	833,615	32%
Religion	11	7%	376,607	15%
Entertainment	3	2%	17,137	1%
Science	2	1%	55,784	2%
Sexuality	1	1%	9,926	0%
TOTAL	169		2,589,376	

When looking at the sub-categories (see Table 4), following those used in literature, we find that paintings often illustrate an article about the art piece itself, but can also serve as portraits (of artists, of historical and political figures, of mythological and religious characters). Paintings are also used to illustrate places. The share of views is noticeably larger for artists' biographies yet views to specific art pieces receive a relative low share of views. Views to articles about historic, political and mythological figures are relatively large in proportion to the number of articles.

Table 4 Number of articles and number of views per subcategory

Primary category	Secondary category	Records	%	Views	%
Arts	Artists	50	30%	585,838	23%
	Art	33	20%	111,365	4%
	Fashion	5	3%	36,417	1%
	Museums	5	3%	112,069	4%
	Literature	2	1%	1,159	0%
	Music	1	1%	8,790	0%
Geography	Culture	18	11%	233,118	9%
	Places	10	6%	206,494	8%
	Countries	1	1%	1,057	0%
History	Historical figures	10	6%	270,815	10%
	Events	3	2%	2,982	0%
	Political figures	2	1%	459,593	18%
	Wars	2	1%	1,542	0%
	Other	9	5%	88,282	3%
	Event	1	1%	10,401	0%

Religion	Mythology	6	4%	278,971	11%
	Religious depictions	2	1%	6,738	0%
	Religious terms	2	1%	37,412	1%
	Other	1	1%	53,486	2%
Entertainment	Sports	1	1%	8,131	0%
	Other	2	1%	9,006	0%
Science	Natural science	1	1%	33,299	1%
	Scientists	1	1%	22,485	1%
Sexuality	Other	1	1%	9,926	0%
	TOTAL	169		2,589,376	

Aggregating the share of articles and views that are directly related to the artist, the painting, or the museum, we find that paintings illustrate over half of the articles in our sample. All other categories have been grouped as non-art related and actually receive two thirds of all views to articles containing a painting from the Metropolitan Museum of Art.

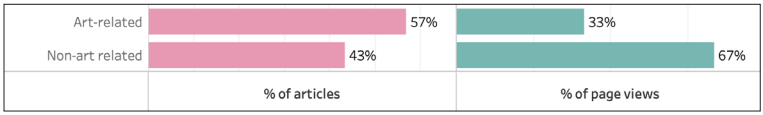

Figure 2 Share of articles and views in art and non-art related categories

Zooming into the most viewed articles we find that the portrait of George Washington and of Christopher Columbus are the most visible portraits in Wikipedia articles, together with the portrait of Vincent van Gogh. Besides the article about the museum, all other articles are the expected general encyclopedic entries.

Table 5 Top 20 most viewed Wikipedia articles including painting images from the Met

No.	Article	Category	Page views	No. paintings
1	George Washington	Non-art related	458,368	1
2	Vincent van Gogh	Art-related	290,658	1
3	Christopher Columbus	Non-art related	247,514	1
4	Metropolitan Museum of Art	Art-related	86,409	4
5	Oedipus complex	Non-art related	83,891	1
6	Sphinx	Non-art related	77,713	1

7	Saint Joseph	Non-art related	53,486	1
8	Amazons	Non-art related	50,834	1
9	Genoa	Non-art related	47,822	1
10	Paul Cézanne	Art-related	43,573	1
11	Scythians	Non-art related	43,225	1
12	Pandora	Non-art related	40,904	1
13	Wall Street	Non-art related	40,622	1
14	Greeks	Non-art related	33,588	1
15	Pet	Non-art related	33,299	1
16	Édouard Manet	Art-related	32,293	2
17	El Greco	Art-related	32,142	1
18	Five Points, Manhattan	Non-art related	31,957	1
19	Eve	Non-art related	29,672	1
20	Eugène Delacroix	Art-related	24,367	2

We further gather data from the Metropolitan Museum of Art website in August 2018. We identified the 119 paintings in our dataset and collected information about the display status, being currently exhibited or not. We found that 52 paintings (or 38.5%) were not on display, hence could not be viewed by visitors of the museum. These paintings, however, did receive nearly 800 views online in one month (see Table 6). The 65 paintings on view at the museum correspond to the better-known paintings, receiving 2.3 million online views in one month.

Table 6 Wikipedia views of paintings (currently exhibited and not on view)

	Not on view	**Exhibited**	**Total**
Number of paintings	47 (39.5%)	72 (60.5%)	119
Wikipedia views May 2018	736,298	1,942,628	2,589,376*
Number of articles*	60	127	

*Articles may include both paintings on view and exhibited. The number in this table shows the number of views to articles that included a painting in each category, hence views to 'not on view' and 'exhibited' cannot add to 100%.

Unsurprisingly, we find a visible long-tail of the usage of paintings in various articles. Twenty-six paintings were used in more than one article, among the most used paintings we find a portrait of El Greco and a portrait representing Christopher Columbus, currently not on view at the museum. The painting by Gustave Moreau entitled *Oedipus and the Sphinx* from 1864 is used in pages about the artist, the year, the sphinx, its collection William H. Herriman, but also on 'Oedipus complex' and 'Phallic stage', increasing its visibility considerably.

Table 7 Top ten most viewed painting images

No.	Painting	Views	N. articles	Exhibited
1	Emanuel Leutze (American, Schwäbisch Gmünd 1816-1868 Washington, D.C.) - Washington Crossing the Delaware - Google Art Project.jpg	458,970	2	Yes
2	Portrait of a Man, Said to be Christopher Columbus.jpg	369,066	9	No
3	Vincent van Gogh - Wheat Field with Cypresses - Google Art Project.jpg	295,771	3	Yes
4	Oedipus and the Sphinx MET DP-14201-023.jpg	179,274	5	Yes
5	El Greco - Portrait of a Man - WGA10554.jpg	146,687	9	Yes
6	Annunciation Triptych (Merode Altarpiece) MET DP273206.jpg	123,049	5	Yes
7	Rosa bonheur horse fair 1835 55.jpg	117,027	6	Yes
8	JoanOfArcLarge.jpeg	86,409	1	Yes
9	Majas on Balcony by follower of Francisco de Goya.jpg	86,409	1	No
10	Eugène Delacroix - Ovide chez les Scythes (1862).jpg	69,144	3	Yes

4 Discussion

In our case study of the Metropolitan Museum of Art, we find a comparable yearly visit rate between onsite and online visitors. The physical museum receives 7 million visits a year, of which 2 million are residents of a city of 8.5 million inhabitants, while the collection online receives a similar 7.2 million visitors a year, from a global population. The museum has further expanded its publication channels to include one of the most popular websites worldwide and in 2017 uploaded 360 thousand images of its collection to Wikimedia Commons. We have focused on the painting collection. In less than a year, 119 paintings have been used to illustrate 169 Wikipedia articles in English receiving over 2 million views per month. 47 of those paintings (or 40%) are not available for view at the museum because they are in storage, while online views continued to flow reaching nearly 750 thousand views per month. This is a significant increase in collection visibility.

Lacking socio-economic demographic data on the physical and online visitors of The Met, we rely on the behaviour of the online visits to Wikipedia articles including images of the painting collection, our dataset. We certainly observe greater consumption when a larger share of the collection is made accessible, as in the case of the 47 paintings not exhibited at the museum but available on Wikipedia,

Figures 3a-c Examples of painting images used in Wikipedia articles: a) *Portrait of the Imperial Bodyguard Zhanyinbao* by unidentified artist (1760). https://en.wikipedia.org/wiki/Bodyguard; b) *Oedipus and the Sphinx* by Gustave Moreau (1864). https://en.wikipedia.org/wiki/Sphinx; c) *Christopher Columbus* by Sebastiano del Piombo (1519). https://en.wikipedia.org/wiki/Christopher_Columbus

e-ISSN 2724-3923

illustrating 60 articles receiving 736 thousand views, supporting the results of Brook (2006). Online consumption displays an exponential increase when collections are used in a Wikipedia article, not necessarily within an art context, which by the mere numbers represents a broadening of the consumer base, including those beyond the socio-economic profile of the museum visitor. In fact, digital consumption of paintings increases when art is seen as information. We find lower consumption to pages about the art piece and greater consumption to biographies of artists (29.6%), historical figures (5.9%), political figures (1.2%), and mythological figures (3.5%). Collections have an important information value as they fulfil an information function, that of illustrating important figures.

The literature on cultural consumption assumes a purely hedonic drive, or utilitarian to serve in class dynamics. Consumer behaviour in Wikipedia reveals a more complex picture where both hedonic and utilitarian drives overlap. One important distinction can be found between the consumer actively seeking cultural content, visiting pages related to arts and culture, and the consumer seeking non-arts and culture related content, visiting pages related to history, geography or religion. Both consumers find images of The Met, the former representing a third of the views. That means that two thirds of consumers in our sample are not actively seeking to view paintings from the museum yet in fact do so.

This has four important implications for understanding digital heritage consumption in Wikipedia. First, Wikipedia serves as highly accessible source of easy-to-understand content, lowering the threshold of required ability to process information. Consumers are not required to formulate an 'art' query but can land on an article illustrated with art. Following Ganzeboom (1982), this would result in increased consumption also of art-related topics and highly intellectual art pieces. The lower capital profiles identified by Jarnes (2015) can be expected to benefit the most from such accidental encounters, as paintings would increase knowledge, be meaningful and comprehensible, and present encyclopedic content. However, not all consumers are seeking 'art'. Wikipedia is made of hyperlinks allowing consumers to encounter cultural content even when lacking proper cultural capital to formulate the right search, by serendipity (Stiller 2012). The article 'Pet', for example, has numerous images including one painting from The Met's collection 'Young Lady with parrot', by Édouard Manet **[fig. 4]**. Similarly, the articles 'Fair' and 'Horse trade' includes the artwork 'The Horse Fair' by Rosa Bonheuron.

Second, a recent user study identified Wikipedia fact-checking to represent consumers with high socio-economic capital (Singer et al. 2017). Besides substantially lowering the barrier of time reported to visit a museum, the consumer sensitive to status and branding identified by Jarnes (2015) can easily benefit from having mobile access

Pets in art

Cider in the United States

From Wikipedia, the free encyclopedia

This article's **tone or style may not reflect the encyclopedic tone used on Wikipedia**. See Wikipedia's guide to writing better articles for suggestions. *(May 2016) (Learn how and when to remove this template message)*

In the United States, the definition of *cider* is usually more broad than in Europe. There are two types: one being the traditional fermented product, called *hard cider*, and the second *sweet* or *soft cider*. However, in some regions, *cider* is the alcoholic version, whether made from apples or pears, and *apple cider* is the non-alcoholic version.

Contents [hide]

American Hard Cider in a Bottle

Hard cider [edit]

Cider Making, painting by William Sidney Mount, 1840–1841

The history of cider in the United States is very closely tied to the history of apple growing in the country. Most of the 17th- and 18th-century emigrants to America from the British Isles drank hard cider and its variants. Apples were one of the earliest known crops in the English-speaking New World; ships' manifests show young saplings being carefully planted in barrels and many hopeful farmers bringing bags of seed with them, with the first settlers headed to what is now the Southeast. Within thirty-five years of the settlement of Jamestown in 1607, the land was put to the plow to grow tobacco which provided a source of revenue for the colonists and made British settlement a success in the New World after several failed attempts. However, other edible cash crops were planted, like rice, maize, and apples, since such would have had value in the markets of growing cities like London, Edinburgh, Dublin, and Cardiff.

The earliest known provision for cider making is believed to have been carried on the Mayflower itself in 1620. Halfway through the journey, the ship was caught in a storm and one of its beams cracked badly enough to warrant the consideration of turning back to England. "The great iron screw", taken from a cider press, helped brace the beam to keep the ship from breaking up and did it long enough to make it to the New World.[1] Nine days after the Puritans landed (and perhaps in great thanks for having survived the journey at all) a man by the name of William Blackstone planted the first apple trees in the New England colonies.[2] The first recorded shipment of

Figure 4 Screenshot of the Wikipedia article 'Pet' which contains a section called 'Pets in art' (https://en.wikipedia.org/wiki/Pet#Pets_in_art)

Figure 5 Screenshot of a Wikipedia article including a painting from The Met's collection (https://en.wikipedia.org/wiki/Cider_in_the_United_States)

to information such as 'quick facts' during social encounters, enriching their social currency as proposed by DiMaggio (1987). In this way, fact-checking can be expected to particularly benefit status seekers who in turn gain rich cultural information. The article "Cider in the United States" including a painting by William Sidney Mount in figure 5 serves as example **[fig. 5]**.

Third, not very well know paintings can illustrate a variety of articles, as the Imperial Bodyguard portrayed above illustrating "Bodyguard" is also found in the articles "Dao (sword)", "Chinese archery", "Tungusic peoples", "Qing official headwear", and "Wu Quanyou". The range of topics would reach a variety of consumers, as identified by Jarnes (2015), so that each consumer profile may have the chance to encounter an art piece and to gain utility from encountering something new, from being entertained, from learning or from the easing of the access. The variety of topics available in the encyclopedia will certainly tap into the topical interest of a greater number of visitors, lowering an important barrier to museum visits.

Last, the fact that well-known established superstar museums, including the Metropolitan Museum of Art, are collaborating with Wikipedia points to a shift in the field. Museums are establishing a new tasteful form of cultural consumption (Meyer 2000), that of paintings in the online encyclopedia, by positioning authentic images for public online use. Paintings, and metadata about collections, are available with a CC0 (1.0) license inviting further reuse. Authenticity of cultural consumption can clearly be differentiated from authenticity of digital cultural consumption, where the later guarantees an authoritative source (a museum) that can serve as art and as information.

One may argue that viewing a painting in a Wikipedia article may not take more than a few seconds, if at all. Future research is needed to determine the actual viewing of paintings on such websites – current tools are still rudimentary. However, previous ethnographic research on museum visitors reveals that visitors may miss a significant part of collections, especially in large museums, and that viewing paintings may take one up to 40 seconds, while texts are hardly ever read (for an overview of the literature see Beer 1987). Museum visits are hence but an indication of a view to a painting in display, just like views to Wikipedia articles may indicate the rate of digital consumption. Because of the great pervasiveness of non-art related consumption of paintings (67%), we argue that paintings in Wikipedia can be expected to increase the visual vocabulary of visitors and serve to increase consumer utility during future museum visits. Regarding the art-related views (33%), we believe greater information about the painting, the artist, and even the hosting institution, may lead to greater interest to visit the collections. Though the open data movement and publication of museum collections is relatively new to observe clear social effects, this may prove to be an interesting line of future inquiry.

5 Conclusions

In an increasingly digital world, cultural consumption must be freed from museum websites. We looked at paintings from the Metropolitan Museum of Art available as illustrations of English articles in Wikipedia and analysed their visibility during the month of May 2018. We find that paintings are mostly viewed as illustrations of places (17%) followed by illustrations of important historical events and personalities (16%). When analysing the content of the articles, we find that paintings illustrate art-related content (67%) as well as non-art related content, the later receiving two thirds of total views of our dataset (2 million monthly views). Our results have four main implications to advance the understanding of digital cultural consumption. First, there are accidental consumers who benefit from serendipitous consumption to encounter paintings in an environment requiring low cultural capital and information processing abilities. Second, Wikipedia facilitates mobile fact-checking, which can enrich status seekers' conversation and lower the time barrier to cultural consumption. Third, paintings can be used in a variety of articles yielding utility to a variety of consumers, where the intellectual consumers may gain greater content from the layered visual information provided in a painting. And last, museums are participating in new forms of content distribution that expands tasteful consumption to include a digital realm, by providing authentic images for public use. In this sense, authentic digital cultural consumption may rely on a clear source (museum) as well as quality images in art as well as non-art environments.

Based on our results, we can safely say that digital dissemination of collections in Wikipedia will have an impact in future cultural consumption. Our case study receives 7 million physical visits per year while the paintings in Wikipedia received 2.5 million visits in one month. From the 7 billion inhabitants in the planet, we cannot expect everybody will be able to visit the museum in New York City (it will take one thousand years!). Wikipedia appears as a viable channel to disseminate collections to increase consumer reach, even when the object is in storage or when the museum is physically unreachable to all socio-economic groups (Brook 2016). Publication of collections within the institutional website may very well echo a museum consumer profile yet broadening the publication channels would increase consumer diversity. In Wikipedia, paintings serve as illustrations to art as well as non-art related topics, lowering the threshold to view art, favouring what Stiller (2012) suggests for consumers unable to formulate an art-related query. The use of digital authentic images, provided by museums as representations of their collections, can contribute to what Meyer (2000) identified as institutions contextualising the collective definition of taste. It is up to the muse-

um to approve the use of its collection to illustrate the online encyclopaedia and in so doing to transform the social perception of taste, and our concept of cultural consumption. Viewing images of paintings as illustrations of an online encyclopaedia is not comparable to the viewing of a painting inside a museum, where the context is essentially different. However, the attention spam per painting in museums, if the visitor views the painting at all, is generally well below 40 seconds and hardly includes reading text (Beer 1987). In the online encyclopaedia, the focus to reading may predispose the consumer to view images differently. Future research is needed to better understand the interaction with the paintings in Wikipedia as well as the awareness of consumers regarding the viewing of art.

Lines of future work in fact are many, since the online environment provides new forms of cultural consumption. Certainly, comparing the use of paintings in traditional paper-based encyclopedias may signal whether there is a shift in the use of images and whether this shift can be positioned in relation to the growth of our predominantly image-based communicating culture. The Wikipedia Foundation environment is a fertile ground for observing the behaviour of readers, editors, and content contributors (such as museums) but also for gaining some form of harmonization in the data from various sources, always welcomed by digital humanities researchers. For policy makers it would be interesting to measure the extent to which such wider availability of images leads to different type of research. Further, we only looked at the use of paintings in actual articles, but the paintings are used for a number of other uses, including in user profiles, portals, or other Wiki-related pages.

Our study serves as stepping stone in the understanding of digital cultural consumption beyond the art context, yet it cannot escape a couple of major limitations. First of all, regarding the selection of the dataset. While other collections beyond paintings at The Met and a larger span of analysis may provide a different relation of art and non-art article usage, we believe the general result of art paintings used in a variety of articles with greater views in non-art related topics is recognisable by many museums. The adoption of museum collections as encyclopaedic illustrations has a long history, as can be seen from the name of the file of the most viewed painting in our dataset including the name of the source (Google Art Project). The Met serves as perfect example of a well-known museum, with healthy onsite and online consumer base, that decided to explore new markets. Having only views to articles resulted in a limited metric but nonetheless a rich indication of the potential information value of museum content. In figure 6, an example of an item from the collection illustrating the use of the pineapple fiber to make luxury fabrics decorated with floral embroidery **[fig. 6]**. Thanks to the increased adoption of digital technology at the exhibition galleries

Figure 6
Image of textile collection from the Metropolitan Museum of Art used in Wikipedia articles. https://en.wikipedia.org/wiki/Pineapple#/media/File:Kerchief_MET_25.132.8.jpg

and to the improved metrics online, also within Wikipedia, future research may look at the actual consumer journey onsite as well as online. A richer dataset may allow for further analysis, combining quantitative and qualitative insights, for instance to inquire about perception and practice of digital heritage consumption outside of an art-related context. As data improves, comparisons between online environments may shed light on the process of viewing an art piece, and the ways in which extended or mix reality can support the art experience remotely.

Bibliography

Ateca-Amestoy, V.; Castiglione, C. (2014). "Live and Digital Engagement with the Visual Arts". Paper presented at *18th International Conference on Cultural Economics*, 24-7 June 2014.

Ateca-Amestoy, V.; Castiglione, C. (2016). "Digital Cultural Audiences. Should Cultural Managers Worry About the Digital Divide?". Paper presented at *7th EWACE Workshop*, 4-5 September 2015.

Axiell (2017). *Digitising Collections: Leveraging Volunteers and Crowdsourcing to Accelerate Digitisation*. Report. Axiell. Available at https://alm.axiell.com/wp-content/uploads/2017/04/DigitisingReport-1.pdf.

Beer, V. (1987). "Great Expectations: Do Museums Know What Visitors Are Doing?". *Curator*, 30(3), 206-15. https://doi.org/10.1111/j.2151-6952.1987.tb00664.x.

Beckert, J.; Aspers, P. (eds) (2011). *The Worth of Goods: Valuation and Pricing in the Economy*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:osobl/9780199594641.001.0001.

Brook, O. (2016). "Spatial Equity and Cultural Participation: How Access Influences Attendance at Museums and Galleries in London". *Cultural Trends*, 25(1), 21-34. https://doi.org/10.1080/09548963.2015.1134098.

Cameron, F.; Kenderdine, S. (eds) (2007). *Theorizing Digital Cultural Heritage*. Cambridge: The MIT Press. https://doi.org/10.7551/mitpress/9780262033534.003.0004

Cameron, F. (2007). "Beyond the Cult of the Replicant – Museums and Historical Digital Objects: Traditional Concerns, New Discourses". Cameron, Kenderdine 2007, 49-76.

DiMaggio, P. (1987). "Classification in Art". *American Sociological Review*, 52(4), 440-55.

Estermann, B. (2013). *Swiss Heritage Institutions in the Internet Era. Results of a Pilot Survey on Open Data and Crowdsourcing*. Zurich: Bern Institute for Applied Sciences.

EU (European Union) (2013). *Cultural access and participation*. Report. Special Eurobarometer 399 / Wave EB79.2. November 2013.

EU (2017a). *Special Eurobarometer 466. Cultural Heritage*. Report 2017.7226.

EU (2017b). *Promoting Access to Culture via Digital Means: Policies and Strategies for Audience Development*. Final report. June 2017. Luxembourg: EU.

Evans, G. (2016). "Participation and Provision in Arts and Culture – Bridging the Divide". *Cultural Trends*, 25(1), 2-20. https://doi.org/10.1080/09548963.2015.1135528.

Evrard, Y.; Krebs, A. (2018). "The Authenticity of the Museum Experience in the Digital Age: The Case of the Louvre". *Journal of Cultural Economics*, 42(3), 253-363. https://doi.org/10.1007/s10824-017-9309-x.

Falk, M.; Katz-Gerro, T. (2015). "Cultural Participation in Europe: Can We Identify Common Determinants?". *Journal of Cultural Economics*, 40(2), 127-62. https://doi.org/10.1007/s10824-015-9242-9.

Frey, B. (2000). *La Economia del Arte*. Barcelona: La Caixa.

Frey, B.; Meier, S. (2006). "The Economics of Museums". Ginsburg, V.A.; Throsby, D. (eds), *Handbook of the Economics of Art and Culture*, vol. 1, 1017-47. Elsevier.

Finnis, J.; Sebastian, C.; Clemens, R. (2011). "How to Evaluate Online Success?". *Let's Get Real 1* series. Report from the Culture24 Action Research Project.

https://www.keepandshare.com/doc/3148918/culture24-how-toevaluateonlinesuccess-2-pdf-september-19-2011-11-15-am-2-5-meg?da=y&dnad=y.

Fyfe, G. (2004). "Reproduction, Cultural Capital and Museums: Aspects of the Culture of Copies". *Museum and Society*, 2(1), 47-67. https://doi.org/10.29311/mas.v2i1.2783.

Ganzeboom, H. (1982). "Explaining Differential Participation in High-cultural Activities. A Confrontation of Information-processing and Status-seeking Theories". Raub, W. (ed.), *Theoretical Models and Empirical Analyses*. Utrecht: E.S. Publications, 186-205.

Hess, C.; Ostrom, E. (eds) (2007). *Understanding Knowledge as a Commons*. Cambridge (MA): MIT Press.

Hiscox (2017). *The Hiscox Online Art Trade Report 2017*. London: Hiscox Ltd.

ITU (International Telecommunications Union) (2017). *ICT Facts and Figures 2017*. Geneva: ITU.

Jarnes, V. (2015). "Modes of Consumption: From 'What" to 'How" in Cultural Stratification Research". *Poetics*, 53, 65-79. http://dx.doi.org/10.1016/j.poetic.2015.08.002

Jullien, N. (2012). "What We Know About Wikipedia: A Review of the Literature Analyzing the Project(s) (May 7, 2012). Available at SSRN: https://ssrn.com/abstract=2053597 or http://dx.doi.org/10.2139/ssrn.2053597.

Kats-Gerro, T. (2004). "Cultural Consumption Research: Review of Methodology, Theory, and Consequence". *International Review of Sociology*, 14(1), 11-29. http://dx.doi.org/10.1080/0390670042000186743.

Mestyan, M.; Yasseri, T.; Kertesz, J. (2013). "Early Prediction of Movie Box Office Success Based on Wikipedia Activity Big Data". *PLoS ONE*, 8(8), e71226. http://dx.doi.org/10.1371/journal.pone.0071226.

Meyer, H.-D. (2000). "Taste Formation in Pluralistic Societies. The Role of Rhetorics and Institutions". *International Sociology*, 15(1), 33-56. https://doi.org/10.1177%2F0268580900015001003.

Miheij, S.; Leguina, A.; Downey, J. (2019). "Culture is Digital: Cultural Participation, Diversity and the Digital Divide". *New Media and Society*. https://doi.org/10.1177/1461444818822816.

Nauta, G.J.; Van den Heuvel, W.; Teunisse, S. (2017). *Europeana DSI 2 – Access to Digital Resources of European Heritage*. D4.4. Report on ENUMERATE Core Survey 4. The Hague: DEN Foundation.

Navarrete, T.; Borowiecki, K. (2016). "Changes in Cultural Consumption: Ethnographic Collections in Wikipedia". *Cultural Trends*, 25(4), 233-48.

NEA (National Endowment for the Arts) (2015). *When Going Gets Tough: Barriers and Motivations Affecting Arts Attendance*. NEA Research Report #59. Washington: National Endowment for the Arts.

Ratkiewicz, J. et al. (2010a). "Traffic in Social Media I: Paths Through Information Networks". *IEEE International Conference on Social Computing*. https://doi.org/10.1109/SocialCom.2010.72.

Ratkiewicz, J. et al. (2010b). "Traffic in Social Media II: Modelling Bursty Popularity". *IEEE International Conference on Social Computing*. https://doi.org/10.1109/SocialCom.2010.63.

Reinoso, A. et al. (2012). "Most Popular Contents Requested by Users in Different Wikipedia Editions". *Proceedings KEOD 2012*. 4th International Con-

ference on Knowledge Engineering and Ontology Development. Barcelona 4-7 October 2012.

Singer, P.; Lemmerich, F.; West, R.; Zia, L.; Wulczyn, W.; Strohmaier, M.; Leskovec, J. (2017). "Why We Read Wikipedia". *2017 International World Wide Web Conference Committee*. 3-7 April 2017, Pert, Australia. ACM 978-1-4503-4913-0/17/04. http://dx.doi.org/10.1145/3038912.3052716.

Spoerri, A. (2007a). "Visualizing the Overlap Between the 100 Most Visited Wikipedia Pages in September 2006 to February 2007". *First Monday*, 12(4). http://firstmonday.org/htbin/ cgiwrap/bin/ojs/index.php/fm/article/view/1764/1644.

Spoerri, A. (2007b). "What is Popular on Wikipedia and Why?". *First Monday*, 12(4). http://www.uic.edu/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/1765/1645.

Stiller, J. (2012). "A Framework for Classifying Interaction in Cultural Heritage Information Systems". *Progress in Cultural Heritage Preservation. Proceedings EUROMED 2012*, 141-6. https://doi.org/10.1260%2F2047-4970.1.0.141.

Tallon, L. (2018). "Creating Access Beyond metmuseum.org: The Met Collection on Wikipedia". https://www.metmuseum.org/blogs/now-at-the-met/2018/open-access-at-the-met-year-one.

The Met (The Metropolitan Museum of Art) (2017). *Annual Report for the year 2016-2017*. New York: The Met.

The Met: Open Access at The Met. https://www.metmuseum.org/about-the-met/policies-and-documents/open-access

Villaespesa, E.; Navarrete, T. (2019). "Museum Collections on Wikipedia: Opening Up to Open Data Initiatives". *MW19: Museums and the Web 2019*. Consulted February 3, 2019. https://mw19.mwconf.org/paper/museum-collections-on-wikipedia-opening-up-to-open-data-initiatives/.

Villaespesa, E. (2017). "Who are the Users of The Met's Online Collection?". Consulted February 20, 2019. https://www.metmuseum.org/blogs/collection-insights/2017/online-collection-user-research.

Walsh, P. (2007). "Rise and Fall of the Post-Photographic Museum: Technology and the Transformation of Art". Cameron, Kenderdine 2007, 19-35.

Weilenmann, A. et al. (2013). "Instagram at the Museum: Communicating the Museum Experience through Social Photo Sharing". *CHI '13: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1843-52. https://doi.org/10.1145/2470654.2466243.

Wu, J.; Lu, X. (2013). "Effects of Extrinsic and intrinsic Motivators on Using Utilitarian, Hedonic, and Dual-Purposed Information Systems: A Meta-Analysis". *Journal of the Association for Information Systems*, 14(3), 153-91. https://doi.org/10.17705/1jais.00325.

Zins, M. (2017). "Fine Arts Museums at a Crossroads: Between Core Mission and Adaptation to New Tourist Clients". *Journal of Marketing Trends*, 4(2), 57-64.

Annex

Table A1 Top ten most viewed painting images (currently not on view)

No.	Painting image	Views	N. articles
1	Portrait of a Man, Said to be Christopher Columbus.jpg	369,066	9
2	Majas on Balcony by follower of Francisco de Goya.jpg	86,409	1
3	Departure of the Amazons MET DP318355.jpg	50,834	1
4	Portrait of the Imperial Bodyguard Zhanyinbao.jpg	36,567	5
5	Angel of the Divine Presence Bringing Eve to Adam (The Creation of Eve- "And She Shall be Called Woman) (recto); Sketch for the same (verso) MET DP805381.jpg	29,672	1
6	Girls Carrying a Canoe, Vaiala in Samoa MET ap1970.120.jpg	23,920	1
7	Baron Alexander von Humboldt (1769-1859) MET DP-1411-001.jpg	22,485	1
8	Atalanta and Meleager MET DP261342.jpg	22,218	1
9	Madame Félix Gallois MET DP359015.jpg	11,407	1
10	無款 清末 京劇一百人物像 冊 絹本-One hundred portraits of Peking opera characters MET DP280076.jpg	8,790	1

e-ISSN 2724-3923

e-ISSN 2724-3923

magazén
Vol. 1 – Num. 2 – December 2020

Neither Literature nor Object: Children's Writings in the Digital Public Realm

Lois Burke
University of Edinburgh, UK

Kathryn Simpson
University of Glasgow, UK

Abstract The interpretation of children's writings has often presented a particular challenge to Galleries, Libraries, Archives and Museums (GLAM), as the represented child has historically been deprived of agency, and children's writings are neither 'literature' nor traditional display objects. In this article we will explore the methodologies of representation that are associated with the merging of children's history and digital humanities. We will lay out an approach for digitally representing children's writings held in museums. We will demonstrate the possibilities that have been put forward by librarians, archivists and curators internationally, and explore the tools and approaches that have emerged from the field of digital humanities for re-presenting the agency of the child creator and the child visitor within memory institutions. Moreover, in this article we will propose that the digital environment facilitates a critical site of experimentation in displaying children's collections that allow creator, object, context, critique, and visitor to be equally valued.

Keywords Museums. GLAM. Digital humanities. Ethics. Digitisation. History of childhood. Child diaries. Youth culture. Visitor experience. Manuscripts. Collections. Hybrid digital spaces.

Edizioni Ca'Foscari

Peer review

Submitted	2020-07-21
Accepted	2020-11-03
Published	2020-12-22

Open access

Citation Burke, L.; Simpson, K. (2020). "Neither Literature nor Object: Children's Writings in the Digital Public Realm". *magazén*, 1(2), 249-270.

DOI 10.30687/mag/2724-3923/2020/02/005

1 Introduction

As GLAM institutions are couched in an ever more networked industry which melds digital and analogue experiences, they must increasingly look to digitise aspects of their collections, for visitors who expect to experience digital exhibition interactions and access objects via digital means. The method of choosing what material is converted into a digital form, and the manner in which this is conducted, is often complicated and inconsistent. In this article we argue for the uniqueness of children as creators of museum objects and museum visitors. In particular, we suggest that the interpretation of children's writings has often presented a particular challenge to GLAM institutions, as the child has historically been deprived of agency, and children's writings are neither 'literature' in a canonical sense nor are they objects which can be easily understood through traditional means of display.[1] The nuances of children's writing collections, and the ways in which they can be explored digitally and ethically, will be the focus of this article.

Pluralistic methodologies of representation are found with the merging of children's history and digital humanities. The digital environment facilitates a critical site of experimentation in displaying children's collections that allow creator, object, context, and visitor to be equally valued. Ultimately, in this article we will explore the theoretical and real-world implications of the digital humanities as a bridging ontology between publics and museums, and offer reflections on and recommendations for meaningful digital engagement with children's collections. We will highlight that establishing a rigorous ethics of collections digitisation is of particular importance when considering those objects which were made by children. The focus will be on 19th and 20th century collections of childhood writings, as this reflects the authors' primary research interests, yet the experiences of contemporary youth in museums is given significant consideration. First, our critical methodology will be established by examining recent developments in the history of childhood, museums, and digital projects, before we consider The Museum of Childhood in Edinburgh as a case study for digital engagement. To finish we gesture to two digital projects, "Girl on a Whaleship" and "The Anne Frank House Museum", which, in our view, have successfully implemented digital mediation of children's writings held in museum collections.

1 Children's writings are subjects of academic interest primarily for children's literature studies, juvenilia studies and life writing studies. Journals associated with these fields tend to interpret children's writings as texts ripe for close reading and other literary analyses, or even as works of art, but their status as objects in collections is not a primary concern in these venues. As children's writings are ambiguously classified as objects by museums (and our case study venue The Museum of Childhood in particular) we will refer to them as objects for the purpose of this article.

e-ISSN 2724-3923

1.1 The Current Field of the History of Childhood

A fascination with liberating children's archival material from obscurity has pervaded recent historical research. In the past decade many conferences and journal articles have hinged on the theme of "uncovering", "liberating" or "rediscovering" archival material related to the child, or alternatively on centralising the child's voice and foregrounding instances of children and young people "speaking up and speaking out".[2] Myriad social and cultural shifts are responsible for this dual focus on archival research and the child's voice.

The power of the child's voice can be seen in contemporary political and activist contexts, which is exemplified in pioneering figures like Malala Yousafzai and Greta Thunberg. The importance of documenting children's lives has been impacted by the phenomenon of a global extended period of adolescence for young people, as well as an increase in youth self-publication, through social media and other cultural products. Many new digital platforms are commanded by youth users; 41% of the 800 million monthly users of TikTok, a mobile video creation platform, are between 16-24 years old (Beer 2019). In June 2020 teenage users of TikTok used their collective influence to digitally disrupt American President Donald Trump's election rally in Tulsa, after registering for thousands of tickets with no intention of attending the event (Lorenz, Browning, Frenkel 2020). This example highlights the self-fashioning and curatorial use of digital media by children to engender agency and create narratives of their own contemporary childhood.

Many scholars working on the history of childhood acknowledge the ongoing methodological challenge that is representing and interpreting children's experiences, while acknowledging the adult intermediary that is always present in this process. In the first volume of the *Journal of Juvenilia Studies* published in 2018, Victoria Ford Smith asked in her article "Exhibiting Children", "What if we begin [...] with the assumption that the child artist is an intentional agentic subject [...]?". Although Ford Smith writes here about child artists, her radical questioning of existing adult assumptions that surround the child artist can equally be applied to the child writer. In the words of the anthropologist Allison James,

2 The Children's History Society UK conference, scheduled for June 2021, is entitled *Children and Youth Speaking Up and Speaking Out*. In 2017 a workshop *Speaking When Spoken To': Re-Integrating the Experiences and Perspectives of Children into Historical Research* was held at the University of Edinburgh, and in 2018 the conference *Opening Up the Archives: Collections, Collaborations, and Forgotten Histories in Children's Literature* was also held there. Recently published articles include: Hoegaerts 2016; Alexander 2012; Moruzi, Smith 2012.

> giving voice to children is not simply or only about letting children speak; it is about exploring the unique contribution to our understanding of and theorising about the social world that children's perspectives can provide. (2017, 262)

Understanding children's voices from the past, and the ways in which youth today engage with these histories, can be part of the opening up current practice in GLAM institutions, to see children as the sites of their own knowledge creation as well as a glimpse into the future of childhood representation and archival practices. Children are not only individualised digital actors but they also represent one of the largest museum visitor groups and thus one of the primary user groups museums can, and do, engage with. For example, "in the United States, about 80% of museums provide educational programs for children (Bowers 2012) and spend more than $2 billion a year on education activities (American Alliance of Museums 2009)" (Andre 2017, 49). Part of the impetus of this current research is to reinforce the ambiguity or abstraction of each unique museum experience and to show how the evolving relationship between physical and digital object is part of this democratising and individualising boundary breaking.

1.2 Representations of Childhood in Museums

Museums define and classify objects "according to the frameworks of knowledge that allow them to be understood" (Hall 1997, 191). The discursive formations which make up the "frameworks of knowledge" by which the museum presents the child are rooted in the authority of the adult to survey childhood in a highly specified way. Therefore, children are often underrepresented in museum exhibitions and collections, despite the universality of the biological experience of childhood. When their lives are interpreted and exhibited for public consumption, there is a risk of an uncomplicated and sentimentalised portrayal of childhood, as an idyllic and distant idea. The gimlet eye of reminiscence is used to curate exhibitions which represent an adult interpretation of a previous time. The "sanctity of a happy childhood" is a universal value, a value which finds itself repeatedly presented in the display of childhood (Hamilton 1997, 119). But the ways in which groups of people are interpreted positions them as subjects, and this presentation needs to be problematised. Such signifying practices produce meaning which "involve relations of power, including the power to define who is included and who is excluded" (Woodward 1997, 15). As with other marginalised groups, children are not seen to be able to be creators of their own worlds. To suggest that narratives of children focus solely around adult notions of child-

play is to deny the agentic abilities of children to narrate their own lived experience. Todorova notes that "letters and diaries by children have the potential to represent a more individual contribution to historical writing from a child perspective" (2017, 25), and we argue that it is digital methodologies which have the capacity to facilitate this participation.[3] The goal of utilising digital means is not to rewrite history nor to suggest that the overarching narrative of historical events is wrong, rather to contextualise history in a way that was previously unachievable. Digital methodologies have the capacity to show that histories can be read against the grain and in doing so provide scope for wider and more inclusive narratives to be told.

It is not only the child as a creator which needs to be renegotiated, but also the child as a visitor. The digital can be an experimental site which enables children to engage with objects and texts they would not usually be allowed to. Described by some as "messy", "leaky", "chaotic" and "undisciplined" museum users, it can be difficult to find ways to facilitate children's access and engagement (Shildrick 1996; Birch 2018; Scollan, Farini 2020). Children do not often fit into how a museum space has been socialised, and their engagement is often looked down upon. Unfortunately, in such spaces "children's embodied practices, and those of their family or guardians, are not always recognised with such open-minded enquiry, often being judged to be out of kilter with institutional mores" (Birch 2018, 519). Huhtamo observed that

> any exhibit with something to click, pull, or rotate drew hands like a magnet, but normally the experience both started and ended there. It was as if there had been nothing at all to be gained beyond the momentary acts of punching and tapping, pushing and pulling. (2017, 65)

We would argue that such a biased and restrictive reading of museum interaction actively marginalises children. Huhtamo goes on to say that the "user interface has become The Thing, instead of serving as a gateway to more cerebral pleasures and discoveries (as I believe it is supposed to do)" (65-6). If the "user interface has become The Thing", then we reason that as a manipulatable interstitial space the user interface already suggests itself to be foundational to facilitating child agency.

Finding ways to structure and display children's works can enable the renegotiation of the social and material context of these objects. As digital tools feature strongly in contemporary methodologies that can be used to circumnavigate the problems of space and physical

3 This is explored further in the case studies in § 3.1.

e-ISSN 2724-3923

display, it follows that they can lead to the re-agentic individualisation of children within museum collections, "discontinuing a long history of othering children through space and activities designed 'for' them" (Birch 2018, 517). Children have difficulty presenting themselves as self-directed creators or users because of the inherent power and politics of a museum space which is why we argue for the capabilities of digital spaces to enable transformative engagement.

The project of recuperating children's history is not divorced from the drive to actively display the history and achievements of other historically marginalised individuals. Mary Jo Maynes recognises that seeking to re-conceptualise children's agency is analogous to recognising the agency of women (2008). As signifying terms, 'women' and 'children' are often homogenised groups considered 'other' to the norms of representation and narrative. Shildrick writes that "the binary structure which characterises Western epistemology is no less entrenched in the ontology of self and other, or in the categories of sameness and difference" (1996, 5). Facilitating child agency in representation and display in a museum has the capacity to challenge those norms and create new forms of knowledge. To return to our opening gambit, the ambiguity of children's writing as juvenilia not considered to have aesthetic merit as literary works and as complicated objects which upon being presented in a display case are stymied of their whole-ness, benefits from non-binary forms of engagement.

Digital tools can enable the actor, event, time, location, and medium to coalesce into a singular space of interpretation and facilitate the visitor's understanding of contextual digital information without prior interpretation of the object. The digital space can be hybrid, coming out of the hegemonic institutional framework of the physical museum and yet malleable, extrapolatable and with the capacity to enable variant readings and interactions – a third space. A third space

> constitutes the discursive conditions of enunciation that ensure that the meaning and symbols of culture have no primordial unity or fixity; that even the same signs can be appropriated, translated, re-historicised and read anew. (Bhabha 2004, 55)

The creation of a new cultural understanding can only exist in a space, a third space, which is neither reliant solely on one or other absolute reading of cultural history, within a museum context we suggest that a hybrid digital space fulfils this role. Bhabha writes that this third space "displaces the histories that constitute it, and sets up new structures of authority, new political initiatives, which are inadequately understood through received wisdom" (Bhabha, Rutherford 1990, 211). This third space accommodates a shifting power dynamic that enables the child author-creator to be heard with minimal adult mediation.

e-ISSN 2724-3923

Many museums aim to represent the history of childhood and display a balance of amusement and instruction which tends to characterise the lives of young people. In the United Kingdom the museums which primarily represent the history of childhood are the National Trust Museum of Childhood, Derbyshire; the Victoria and Albert Museum of Childhood, London; the Museum of Childhood in Edinburgh and the Highland Museum of Childhood in Strathpeffer. In regard to digital engagement, all of these museums have websites; only the National Trust museum has a publicly-searchable collections database, and none of the websites have digitised collections.

2 Digital Work in the History of Childhood

If the reparation of children's histories is akin to the reparation of women's histories (Maynes 2008), then the ethical concerns of representing children's collections digitally are equally as complex. Michelle Marovec (2017) has outlined various issues that are inherent in digitising the creations of marginalised figures. Although many digital projects have begun in the spirit of "techno-optimism" (Moravec 2017, 189), the reality is that most digital projects are at the mercy of short funding windows and often focus on digitising a specific, known aspect of a collection, which in turn can reinforce marginalisation of other individuals represented in a collection. There is also the "bias squared" idea, put forward by Oonagh Murphy of the Museums and AI Network, which recognises that if museum exhibitions and museum stakeholders have biases, then inevitably digital technologies employed in museums will also have biases (Murphy, Villaespesa 2020). Attention must be paid to these digital concerns in the unique context of representing children's perspectives.

Museums' engagements in the digital realm are by no means consistent. Digital projects are affected by opportunity and economy, neither of which are equitable across the sector. In a study by the Network of European Museum Organisations (NEMO) published in 2020, it was found that

> 3 out of 4 museums report that their biggest obstacles regarding the digitisation and online accessibility of the permanent collections are insufficient resources (money) and insufficient (time of) staff.

Furthermore,

> less than 20% of the collections of responding museums are available online. This means that less than half of the digitised objects are available to the public. (NEMO 2020)

Although a commitment to digitisation is a necessary responsibility for most – if not all – museums, in reality this is not currently practicable on a mass scale.

Museums have had a tendency to "overprovide meaning for (adult) visitors rather than concentrate on the object-audience encounter" (Birch 2018, 521). We argue that digitisation brings a primacy to the object-audience encounter which not only reduces the reliance on a proscriptive text-based interpretation of an object but also gives the visitor agency over their own engagement with the object. The possibilities for digital engagement abound, from visitor engagement with object surrogates and narrative responses to objects created and recorded by visitors (Ferris et al. 2004), to exhibits that sense users and project content at them – altering the size of projection based on user distance from the object (Wolf et al. 2015), to "digital augmentation" of tangible museum objects (Not et al. 2019) and generative 3-D object creation of museum objects with open accessibility for printing and remixing (Smithsonian 2020). In fact, certain digital aspects are expected by museum visitors and are becoming universal elements of exhibitions. Carrozzino et al. (2018) acknowledge that a challenge to the creative cultural industry is an increasingly active audience who seeks to partake in the process of communication and interpretation. Touch-screen exhibits have already become ubiquitous to the modern museum experience. Haidy Geismar argues that

> digital screens have become not just the vehicle for delivering information but objects of appreciation in their own right that mimetically appropriate the effects and engagements of glass cases. (2018, 13)

Developments in digital history underpin museums' engagements in the digital realm. Digital history is aligned with the broader field of digital humanities (DH), "wherein computational methods are implemented in pursuit of humanistic questions" (Romein et al. 2020, 293). DH is an interdisciplinary venture which seeks to reveal, better represent and engage with an array of texts from the humanities through digital means. The elasticity of DH means that it can often cater to those texts and projects which have historically proved problematic to analyse and represent. The records of children are notoriously difficult to uncover for researchers. In fact, there currently exists no physical or digital archive dedicated to children's history and culture in the Anglophone world. Although there are many academic or researcher-led digital projects which engage with diaries or children's narrative experiences, these projects are often unfortunately

separate from the museum space.[4] Whether due to funding issues, lack of physical space or technology, or because the source material comes from multiple repositories, rarely do the institutions which these projects have taken their source material from display such projects. This suggests there is still a siloing between forms of discourse which engage with children's writings. As such the available information concerning children's collections is disparate, disjointed and diffuse. The use of DH methods, with their ability to bring together history, culture and scholarship and present research to new audiences in novel ways, has the potential to re-engage and represent the narratives of this marginalised group. DH tools provide ways to circumnavigate the previous lack of access to children's histories, by bringing together small and disparate pieces of information, digitally presenting the physical object, and opening displays to various possibilities for interaction.

We suggest that the merging of DH and children's collections goes further than that. We propose that digital mediation of children's works has the capacity to create a third space within a museum environment; a hybrid space within which to create new notions of identity beyond those which are prescribed. Children are inherently bound to adults as the agents of power and control in their lives, they rarely have autonomous shaping power over how their identity is recorded and classified. This is not to say that adult mediation is damaging; it is necessary and often beneficial. But, it is important that adult mediation facilitates a space within which children's own representation is primary and, to borrow a term from children's literature, exhibitions are open to a "crossover" audience, including both adults and children. To return to our earlier point, to "other" children and see them as "they", is to make them "the subject of verbs in a timeless present tense", and to see them "not as [the result of] a particular historical event but as an instance of a pre-given custom or trait" (Pratt 1985, 120). In other words, the displayed work of a single child comes in to represent the "essence" of all children's creations (Gelman 2003). This essentialist view of childhood should not be the norm in exhibitions.

In engaging digitally with their children's history collections, museums can allow for pluralistic readings which centralise the role and agency of the child-creator. It is this ability to represent children's

4 Projects such as *The Great Diary Project* (https://www.thegreatdiaryproject.co.uk/), *Children's Diaries During the Holocaust* (https://encyclopedia.ushmm.org/content/en/article/childrens-diaries-during-the-holocaust), the BBC's *Childhood and Evacuation in WWII* (https://www.bbc.co.uk/history/ww2peopleswar/categories/c1162/). The British Library's work on Children's writings from WWI is a notable exception (https://www.bl.uk/world-war-one/articles/childrens-experiences-of-world-war-one).

writings alongside mainstream or primary museum resources that is especially powerful, given that this can provide access to children's histories; which, as collection subjects, have been on the margins. As providing stewardship, preservation and access to collections are a curator's core responsibilities, digital techniques must be called upon to provide ongoing engagements with the history of childhood.

2.1 The Museum of Childhood and Digital Engagement

Established in 1955, the Museum of Childhood in Edinburgh was the first museum in the world to specialise in the history of childhood. Their collections of children's writings will be the focus of our recommendations for digitisation and digital engagement. Started by a former Edinburgh city councillor, Patrick Murray, the original purpose of the museum was to tell a history of childhood for adults; it was not to be a museum for children. Murray's eccentric approach to curatorial work can be seen today in outdated albeit humorous remnants of his documentation and labels. His sarcastic and sometimes derisive style of writing labels demonstrates that he projected his personal subjective responses on the objects in the collection, a technique that has become obsolete in recent years. The idea of the Museum of Childhood as a child-free zone has also long been dismissed.

As a free museum that receives more than 200,000 visitors from around the world annually, the Museum of Childhood is a flagship museum of Edinburgh Council's 13 Museums and Galleries venues. The two onsite stores contain collections including toys, games, clothing, objects related to children's education and medicine. A book collection held in a separate store in Edinburgh contains a further estimated 20,000 items, including children's novels, religious books, fairy tales, magazines and annuals. The focus of this article, children's writings, is also held in the Museum's collection.

The main gallery in the Museum underwent substantial renovations in 2017. Mostly funded by a Museums Galleries Scotland grant, the re-imagined gallery included digital elements, including an edited video, oral history listening stations, and a large digital touch-screen photo album. On visiting the museum in November 2019 the Authors noted the positive and persistent engagement with this form of digital interaction. Both adult and child visitors cooperatively engaged with the digital touch-screen photo album in particular. They talked together and created narratives which extended beyond the contextual digital information given about the images of the children, such as the activity they were doing, or related their own experience to the geographical location from which the image was taken (Burke, Simpson 2019). This vignette evidences visitors' willingness to not only interact with digital displays as a novel form of museum inter-

action but also how visitors cite their own lived experience in the wider socio-cultural environment. This new initiative presented collections in novel ways, and represented collections which had never been exhibited before. Moreover the digital elements involved significant collaboration with individuals and institutions: the video clips were sourced from the National Library of Scotland's Moving Image Gallery, and the curators invited staff from across Edinburgh City Council to submit their home photos to be included in a digital album.

The Museum's rudimentary engagement with the digital realm invites speculation on how this engagement can be further developed, in line with the view that children's perspectives still need to be better heard. The digital space offers the opportunity to rethink how we choose to engage with our own socio-cultural history; foundational to rethinking that history is bringing in a multitude of perspectives that have always been there but have not necessarily been heard or seen, as with the digital photo album described above.

2.2 The Museum's Collections and the Child's Perspective

The elusive and complicated status of the child's perspective is arguably epitomised in children's writings. As Moruzi, Musgrove and Pascoe Leahy state in the introduction to their 2019 edited collection *Children's Voices from the Past*,

> finding children's voices remains methodologically challenging and theoretically complex, but the ethical imperative of the task demands that historians continue in the attempt. (20)

There are multiple considerations surrounding accurate and ethical interpretation of these objects. For example, scholars have identified ethical issues with the framing of Anne Frank's diary, one of the best-known examples of children's writing. Many have written on the 'misuse' and 'appropriation' of Frank's diary, particularly regarding the editing of Frank's words, and the book covers of the various editions of the text. As Todorova notes about children's writing more generally, the

> the adult agents involved in the production of the book for mass consumption, however well intentioned, ultimately are silencing – or at least muting – the child-author's voice by speaking for and about her, by translating and interpreting her rather than enabling child and adult readers to hear her. (2017, 26)

The postwar sensibilities of those publishers who interacted with Frank's diary obscured her original intentions. This meant that, in

the words of Ozick, "the diarist's dread came to be described as hope, her terror as courage, her prayers of despair as inspiring" (1997, 22). A misuse or unethical appropriation of children's writings is something that must be avoided, in exhibitions as well as in publication.

The types of writing in children's documentary collections vary widely, from diaries to letters and periodical magazines, and at the Museum of Childhood they have been stored in boxes entitled "Communication" and "Creative Writing". The Museum holds various children's writings from the 18th to the 20th century. These include letters from the year 1770 written by a school girl Isobel Wilson to her mother (accession number 24119);[5] the *Pierrot* magazine (MC86.86) which was written by children between 1911-1915, the contents of which reflects the outbreak of the First World War; and a diary written in 1960 kept by an Edinburgh girl (MC6704). These documents are rich historical resources, as they provide both a written and material historical record. Recent research by Burke (2019), Gleadle (2018; 2019), Sloan (2017) and Pooley (2015) has emphasised the historical value of conducting archival research into collections of children's writings, and drawn attention to the range and variation of these materials. Children's written (documentary) evidence can also offer insights into aspects of youth which are perhaps taboo, for example girls' first-hand experiences of puberty and menstruation can perhaps only be found in few diaries, and no other primary sources (Brumberg 1997).

Although children's writings can be visually appealing to some museum visitors (one part of the child-made magazine the *St. Bernard's Budget* – accession number MC808.96 – from 1892 is on display in the Museum of Childhood's Gallery 4), they present a range of interpretation difficulties for curators. A displayed child-written manuscript is neither a ludic object which is compelling to very young visitors; nor is it 'literature' written by a recognisable author whose name attracts visitors.[6] The manuscript might present palaeographical issues for visitors, and it will be vulnerable to light and other environmental damage. Furthermore, displaying a two-page spread in a bound manuscript volume might not represent the heterogeneity of the text as a whole. Essentially, the manuscript becomes an object which incurs cultural visibility but not readability.

5 *Letter to my Dear Mamma from Isobel Wilson* (1770). The Museum of Childhood, Edinburgh. 24119.

6 Exceptions to this might be juvenila written by authors who became notable in adulthood. Two digital examples of 19th century juvenilia are Virginia Woolf's childhood magazine *Hyde Park Gate News*, and Lewis Carroll's *The Rectory Magazine*. https://www.bl.uk/collection-items/hyde-park-gate-news-a-magazine-by-virginia-woolf-and-vanessa-bell; https://hrc.contentdm.oclc.org/digital/collection/p15878coll30.

A particularly ambiguous and complicated example of children's writing held in the Museum is the *Evergreen Chain* (MC2018.059), a home-made (manuscript) magazine which was written, edited and circulated by a group of adolescent girls living around the UK in the 1880s and 1890s. The girls would contribute poems, short stories, puzzles, and drawings to the editor, who then circulated the bound volume around the group. The volumes of the *Evergreen Chain* are sophisticated documents – research into them has revealed insights into the collaborative nature of youth writings in the late nineteenth century. They are also dynamic: the goals of the magazine changed as the writers aged, and submissions to the writing competition had to be divided into two age categories reflecting the diversity of the younger and older adolescent writers. The feedback on submissions also became more substantial, and the Headmistress of a girls' school was brought in as a critic.

These findings into a unique historic children's culture could only be located through dedicated study, and this information would otherwise be obscured from public knowledge. Presenting the *Evergreen Chain* in a digital format would enable wider engagement and appreciation of these documents which would otherwise require close analysis, and could not be entirely comprehended through traditional means of exhibition.

3 Digital Ethics and Engagement Possibilities with Children's Writings

There are various ways in which children's writings can be presented digitally to allow engagement for adult and child visitors and researchers alike, such as techniques that generate data from the collections and make them machine readable and searchable. These techniques are often primary in providing digital surrogates, or copies, of child-created texts.[7] There is the potential to conduct network mapping using children's correspondence and other shared writings, such as collaboratively-written creative works.[8] Rich metadata is often contained in museum documentation as well as in the textual ob-

7 XML encoding is frequently used to create digital documents that can not only mimic the original physical document but can be more than a digital surrogate being enriched with tooltips, linked to images of the original document and to other contextual documents or objects. An example of the possibilities of XML encoding is given further on with "Girl on a Whaleship".

8 Utilising network mapping digital tools, such as Gephi (`https://gephi.org/`) or GraphCommons (`https://graphcommons.com/`), can demonstrate children's cultural networks and can create a compelling and nuanced argument for children's agency and self-made culture.

jects. In the case of the children's manuscript magazine *Evergreen Chain*, we might know the names of the contributors, their addresses, friends and relations. For example, an adult critic of the *Evergreen Chain* was Mrs A.M. Hitchcock, a Headmistress who influenced the suffragette Emily Wilding Davison.[9] Network mapping can situate children's connections and cultures in a broader historical context which is visually meaningful. Linguistic text analysis platforms can reveal new information about children's writings.[10] The range of information gleaned from this analysis can give critical weight to child writers as creators/authors, and improves understanding of a group or individual creator's literary themes, influences, and levels of literacy. Similar work is done by the Anne Frank House museum website in which Anne's diary is compared to those of other child writers of the Holocaust.[11] Yet, if museums are to facilitate digital engagement with their content and use digital tools to explore that content, how do they evidence ethical practice when they move beyond traditional practices of curation, display and study to the digital re-presentation, re-mixing and exploration? Underneath this umbrella question are a number of factors which we will elaborate on now.

Firstly, as adults in a position of power, museum stakeholders must be comfortable with what is being asked of the content and the user. They then must consider what creator and visitor or user data is required, where it is being presented or displayed, and who, within the institution, is responsible for that data. Museum stakeholders must also consider who can view the data within an institution. The proprietary status of the software, interface or display must be examined. If it is proprietary, then the long term sustainability of the product and an end-of-life data disposal plan must be considered. Similar considerations arise with open source software and hardwares. It is notable that one of the primary issues with digital engagement within a museum environment is the maintenance of digital displays – for example, institutions being unable to get basic items such as touch screens repaired due to being locked into maintenance contracts (Field Notes 2019). If a digital display device requires internet con-

9 See the history of the Headmistress's connection to the suffragette in this history of Kensington Preparatory School, https://issuu.com/kenprep/docs/kps-wrap-aw__our_history_final_artw.

10 Linguistic text analysis platforms such as Voyant Tools (https://voyant-tools.org/) which is a web-based reading and analysis environment for digital texts. It calculates a summary of vocabulary density, average words per sentence, and the most frequently used words in the corpus. It also visualises this data through a word cloud which presents the most frequently used words. Trends shows a line graph of the relative frequencies across the corpus, with a search box.

11 https://www.annefrank.org/en/anne-frank/go-in-depth/holocaust-diaries-anne-frank-and-other-young-writer/.

nection, is endpoint security being used and is there an institutional policy to change default passwords? Finally, there is the question of who owns the data, whether the data actually needs to be collected, and if the amount of data being collected is appropriate in relation to the task at hand. The aim of listing these considerations is not to place a further undue burden of responsibility on the institution that looks to extend its digital engagement, but to make manifest important conditions around the creation of digital content. DH tools only work successfully as a bridging ontology between the physical and digital production of content if they are consistently scrutinised to ensure open and ethical practice. In the following section we will examine two museum projects which have successfully digitally represented children's writings.

3.1 Case Studies of Digital Children's Writings

The website "Girl on a Whaleship" (`http://www.girlonawhaleship.org/`) is an excellent, although dated, example of the digital preservation and curation of children's writings. The site was produced as an online exhibition by Martha's Vineyard Museum in 2010; the building of the site was facilitated by a National Endowment for the Humanities programme, "We The People". The exhibition is built around Laura Jernegan's journal, which is held in the museum's permanent collections.

Laura's journal documents a 3-year whaling voyage she took with her parents, brother, and ship's crew in 1868. On the website one can access the colour scan of the journal, its transcription, and an audio recording of the journal. This tripartite model of presentation ensures maximum accessibility for visitors to the object, and allows new understandings which cannot be sought through accessing the physical manuscript only. Importantly, the wealth of digital contextualising information renders this piece of nineteenth-century children's writing comprehensible to current school-age children. The website provides rich additional interpretation to this already exceptional object. Interactives include a timeline, a history of whaling, a picture gallery, glossary of terms, and an interactive figure of a whaling ship. The picture gallery contains paintings, drawings and objects from Martha's Vineyard Museum's wider collections, which enable the user to understand how the manuscript fits within the museum's larger collection and also relates the manuscript to its historical socio-cultural context.

In the transcribed journal entries, hyperlinks are attached to esoteric vocabulary – that which is either specific to whaling or now obsolete. The digital diary entries provide a diplomatic transcription of the manuscript; they reproduce Laura's grammar, syntax and pres-

entation of her journal as close to the original as possible, including redacted spelling errors indicated by the use of strikethrough. This all contributes to the impression that the transcription work honours the original intention of the author, and in doing so gives credence to the girl author's intent and perspective. It facilitates the autonomy of the child user to learn about whaling history, while simultaneously respecting the original child-created object. "Girl on a Whaleship" is a highly accomplished digital project with respect to the goals of utilising a digital environment to create novel opportunities for engagement with the history of childhood. It stands out as a model of best practice for the representation of children's writings which has paid due diligence to the author, the subject of the writings, the context in which they were written, and also the user who accesses these pages digitally. It allows for full access and analysis of the diary as an historical document.

Another example of accessible and sensitively presented child writings in a digital format can be seen in the sections of the Anne Frank House museum website (`https://www.annefrank.org/en/anne-frank/diary/`) dedicated to her diary. The pages of the website are multifaceted, and display images of Anne's manuscripts and published works, specially commissioned videos, as well as text. These carefully curated web pages have the effect of dispelling the mythology, as we noted earlier, surrounding Anne's written works and allowing the visitor to access the various incarnations of them, which underwent various edits in both manuscript and published forms.

Perhaps the best-known child writer there is, Anne Frank kept a diary while in hiding in a secret annex in Amsterdam during the Holocaust. She wished for her diary to be re-written and published,[12] and after she died in Bergen-Belsen internment camp in 1945, her father Otto Frank sought publication for it. One page on the Anne Frank House museum website dedicated to her diary is entitled "The Complete Works of Anne Frank". The text on this page is set up in a question and answer format, with questions such as "When does Anne get her diary? When does Anne start writing? In which language does Anne write?" followed by succinct responses. This dialogue style seems to represent visitors' frequently asked questions which would be addressed in museum labels in a physical display.

In answer to the question "Does Anne only write in her diary?", we learn about four different types of text that Anne wrote beyond her diary. This is illuminated by the title of the web page – "The Complete Works of Anne Frank" – which is usually reserved for the publications of canonical authors, and has the impact of suggesting that Anne was a child writer who wrote, to borrow the words of Ford Smith, "with

12 Evidenced by her own ongoing editing of her journal and creation of 'clean copy'.

deliberation and insight" (2018, 68) in several different genres. The website also contains pages dedicated to Holocaust diaries written by other young writers. These features facilitate both adult and juvenile learning about Anne Frank's life and works and encourage further research by signalling other child writers and including reference lists for further reading and teaching materials aimed at primary school pupils. This online exhibition of Frank's writings occupies the third space of Pratt's notion of a "contact zone" (1985), in which the child writer is represented ethically, and the present visitor is engaged in a dialogue with the agentic child of the past.

The website also features the Anne Frank video diary, a 15-episode series recorded in Dutch with subtitles in nine languages.[13] The series is recorded as if it is a home-video taken from Anne's perspective. Although anachronistic, this method of filming prioritises the character of Anne's perspective, and it resonates with contemporary forms of childhood self-representation, which is often facilitated through smartphone use. Therefore, this video series achieves the goal of reconciling children's historical status as writers and creators with children's contemporary digital engagements, which we wish to champion in this article.

These examples evidence a unique way of encountering child-written texts held in museums. Both websites facilitate an encounter with these objects in a way which is special, individual, responsive and subjective. They break away from the notion of exhibiting which instruct visitors how to appropriately engage with them, whilst facilitating the encounter with the complexity of the writings.

3.2 Looking Ahead

This article has offered both ethical and practical recommendations for the interpretation of children's writings through the use of digital environments, while advocating for the literary richness they contain. As there is currently no Anglophone digital archive dedicated to children's collections as far as we are aware, we argue strongly for the potential of a large-scale multi-institution digital archiving project.

The critical recommendations presented in this article have overwhelmingly supported the idea that digital means can interpret children's writing to an extent which facilitates engagement, but does not digest them so thoroughly that these complex objects are understood in a single, uncontested way. If "museum objects are said to function as active producers of meaning", as we have argued, then museums must provide visitor experiences that are ambiguous and open (Light

13 https://www.annefrank.org/en/museum/web-and-digital/video-diary/.

et al. 2018, 408). The goal of digital engagement is that it opens up space for engagement between the child created artefact, adult and child visitors, researchers, and curators alike. Yet caution must be taken in the ethics of this digital space; sensitive interpretation decisions, an understanding of the implications of dealing with a marginalised or usually muted group and close adherence to copyright regulations and data legislation are requirements in this configuration.

As creators, visitors and users, children are generally less resistant than adults to engaging with digital interfaces. The creation of hybrid digital spaces can take advantage of this willingness and engender a kind of agency for children that does not present them based on their dissimilarity to adults. Hybrid digital spaces can present an ambiguous in-between space in which the child as creator and consumer might fashion their own narrative. Continuing to exhibit those objects that we currently have greater access to means that museum stakeholders are producing and reproducing bias and distorted arguments, whereas access to a greater digital corpus, both within museums and remotely, means that people can use, reuse, re-interpret and re-present previous binary heteronormative and patriarchal interpretations of children's history. The authors' work on digital archives of Anglophone 19th century women (Ball, Burke, Simpson, forthcoming) has shown a huge disparity in uniformity and presenting information in apparently similar web projects. Our view is that the collections of historically marginalised groups should be as open and accessible as possible.

Museums have a continuing responsibility to act as stewards of their collections; to both preserve them and also allow access to them for as long as access is sustainable. These priorities remain in the digital age, and museums' engagements in the digital realm broadens access to collections in ways that were previously unimaginable. Representing collections that evidence childhood is a crucial yet heretofore overlooked aspect of museums' digital engagements. Our goal has been to explicitly merge insights from museum studies, digital humanities, children's literature studies, and the history of childhood to argue for new critical approaches to children's writings held in museums. After all, children's manuscript writings are strange, uncategorizable texts or objects that are interesting and analyzable from a number of disciplinary perspectives. Future work in this area should seek to model these insights.[14]

14 The Authors are currently developing a digital project which analyses meaning and semiotics in the writing of the multiple child-contributors of the *Evergreen Chain* and will use the outcomes to argue for a stronger research impetus into the historical creative literary works of children.

The hybridity of our histories is an unassailable fact of our lived experience. In a hybrid digital space of engagement, the museum has the possibility to be more than a mainstream cultural institution and to become the site of insurgent counter-hegemonic digital presences. In these spaces, narratives drive learning, the object (explicitly not its interpreters or translators) is centred, and the quality of the digital contextual information explodes dominant knowledge frameworks.

Bibliography

Alexander, K. (2012). "Can the Girl Guide Speak? The Perils and Pleasures of Looking for Children's Voices in Archival Research". *Jeunesse: Young People, Texts, Cultures*, 4(1), 132-45. https://doi.org/10.1353/jeu.2012.0007.

Andre, L.; Durksen, T.; Volman, M.L. (2017). "Museums as Avenues of Learning for Children: a Decade of Research". *Learning Environments Research*, 20, 47-76. https://doi-org.ezproxy.lib.gla.ac.uk/10.1007/s10984-016-9222-9.

Ball, H; Burke, L.; Simpson, K. (forthcoming). *No Discrete Collection: Locating Digital Archives of Nineteenth-Century Women*.

Beer, C. (2019). "Is TikTok Setting the Scene for Music on Social Media?" Globalwebindex. https://blog.globalwebindex.com/trends/tiktok-music-social-media/.

Bhabha, H. K.; Rutherford, Jonathan (1990). "The Third Space Interview with Homi Bhabha". *Identity: Community, Culture, Difference*. London: Lawrence and Wishart, 207-21.

Bhabha, H.K. (2004). *The Location of Culture*. Abingdon: Routledge Classics.

Bhabha, H.K. (2015). "'The Beginning of Their Real Enunciation': Stuart Hall and the Work of Culture". *Critical Inquiry*, 42(1), 1-30. https://doi.org/10.1086/682994.

Birch, J. (2018). "Museum Spaces and Experiences for Children – Ambiguity and Uncertainty in Defining the Space, the Child and the Experience". *Children's Geographies*, 16(5), 516-28. https://doi.org/10.1080/14733285.2018.1447088.

Birch, J.; Parnell, R.; Patsarika, M.; Šorn, M. (2017). "Creativity, Play and Transgression: Children Transforming Spatial Design". *CoDesign*, 13(4), 245-60. https://doi.org/10.1080/15710882.2016.1169300.

Enid Blyton Diary (1960). The Museum of Childhood, Edinburgh. MC6704.

Brumberg, J.J. (1997). *The Body Project: An Intimate History of American Girls*. New York: Vintage Books.

Burke, L. (2019). "'Meantime, It is Quite Well to Write': Adolescent Writing and Victorian Literary Culture in Girls' Manuscript Magazines". *Victorian Periodicals Review*, 52(4), 719-48. https://doi.org/10.1353/vpr.2019.0052.

Burke, L; Simpson, K (2019). *Field Notes. Visitor Observation Main Gallery, 17th November 2020, 13:00-14:30*. Museum of Childhood. Edinburgh.

Carrozzino, M.; Colombo, M.; Tecchia, F.; Evangelista, C.; Bergamasco, M. (2018). "Comparing Different Storytelling Approaches for Virtual Guides in Digital Immersive Museums". De Paolis, L.T.; Bourdot, P. (eds), *Augmented Reality, Virtual Reality, and Computer Graphics*. Cham: Springer, 292-302. https://doi.org/10.1007/978-3-319-95282-6_22.

Choi, Hee-soo; Kim, Sang-heon (2016). "A Content Service Deployment Plan for Metaverse Museum Exhibitions – Centering on the Combination of Beacons and HMDs". *International Journal of Information Management*, 37(1, B), 1519-27. https://doi.org/10.1016/j.ijinfomgt.2016.04.017.

Ciolfi, L.; Bannon, L.J. (2002). "Designing Interactive Museum Exhibits: Enhancing Visitor Curiosity Through Augmented Artifacts". *Proceedings of the European Conference on Cognitive Ergonomics* (European Association of Cognitive Ergonomics, Catania, September 2002), 311-17.

Deleuze, G.; Wolfe, C. (1997). "The Actual and the Virtual". *ANY: Architecture New York*, 19/20, 19.6-19.7. https://www.jstor.org/stable/45048853.

Ferris, K.; Bannon, L.; Ciolfi, L.; Gallagher, P.; Hall, T.; Lennon, M. (2004). "Shaping Experiences in the Hunt Museum: A Design Case Study". *Proceedings of the 5th Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques (DIS'04)*, 205-14. https://doi.org/10.1145/1013115.1013144.

Ford Smith, V. (2018). "Exhibiting Children: The Young Artist as Construct and Creator". *Journal of Juvenilia Studies* 1, 62-81. https://doi.org/10.29173/jjs101.

Geismar, H. (2018). *Museum Object Lessons for the Digital Age*. London: University College London Press.

Gelman, S.A. (2003). *The Essential Child: Origins of Essentialism in Everyday Thought*. Oxford: Oxford University Press.

Gleadle, K. (2018). "Silence, Dissent, and Affective Relations in the Juvenile Diaries of Eva Knatchbull-Hugessen (1861-1895)". *19: Interdisciplinary Studies in the Long Nineteenth Century*. https://doi.org/10.16995/ntn.808.

Gleadle, K. (2019). "Magazine Culture, Girlhood Communities, and Educational Reform in Late Victorian Britain". *The English Historical Review*, 134(570), 1169-95. https://doi.org/10.1093/ehr/cez291.

Gleason, M. (2016). "Avoiding the Agency Trap: Caveats for Historians of Children, Youth and Education". *History of Education*, 45(4), 446-59. https://doi.org/10.1080/0046760X.2016.1177121.

Hall, S. (ed.) (1997). *Representation: Cultural Representations and Signifying Practices*. London: Sage.

Hamilton, P. (1997). "Representing the Social: France and Frenchness in Post-War Humanist Photography". Hall 1997, 75-150.

Hoegaerts, J. (2016). "Recording the Subaltern's Speech. Children's Voices in the Antwerp School Archives, ca 1850-1900". *Journal of Belgian History*, XLVI(1), 62-83.

Huhtamo, E. (2017). "Museums, Interactivity, and the Tasks of 'Exhibition Anthropology'". Grau, O.; Coones, W.; Rühse, V. (eds), *Museum and Archive on the Move: Changing Cultural Institutions in the Digital Era*. Berlin: De Gruyter, 65-82. https://doi.org/10.1515/9783110529630-005.

Ibrahim, N. (2016). "Enacting Identities: Children's Narratives on Person, Place and Experience in Fixed and Hybrid Spaces". *Education Inquiry*, 7(1), 69-91. https://doi.org/10.3402/edui.v7.27595.

James, A. (2017). "Giving Voice to Children's Voices: Practices and Problems, Pitfalls and Potentials". *American Anthropologist*, New Series, 109(2), 261-272. https://doi.org/10.1525/aa.2007.109.2.261.

Light, B.; Bagnall, G.; Crawford, G.; Gosling, V. (2018). "The Material Role of Digital Media in Connecting With, Within and Beyond Museums". *Convergence*, 24(4), 407-23. https://doi.org/10.1177/1354856516678587.

Lorenz, T.; Browning, K.; Frenkel, S. (2020). "TikTok Teens and K-Pop Stans Say They Sank Trump Rally". *The New York Times*. https://www.nytimes.com/2020/06/21/style/tiktok-trump-rally-tulsa.html.

Martha's Vineyard Museum (2010). *Girl on a Whaleship*. http://www.girlonawhaleship.org/.

Maynes, M.J. (2008). "Age as a Category of Historical Analysis: History, Agency, and Narratives of Childhood". *The Journal of the History of Childhood and Youth*, 1(1), 114-24. https://doi.org/10.1353/hcy.2008.0001.

Moravec, M. (2017). "Feminist Research Practices and Digital Archives". *Australian Feminist Studies*, 32(91-92), 186-201. https://doi.org/10.1080/08164649.2017.1357006.

Moruzi, K.; Musgrove, N.; Pascoe Leahy, C. (eds) (2019). *Children's Voices from the Past: New Historical and Interdisciplinary Perspectives*. Cham: Springer. https://doi.org/10.1007/978-3-030-11896-9.

Moruzi, K.; Smith, M.J. (2012). "Colonial Girls' Literature and the Politics of Archives in the Digital Age". *Papers: Explorations into Children's Literature*, 22(1), 33-42.

Murphy, O.; Villaespesa, E. (2020). *The Museums and AI Network – AI: A Museum Planning Toolkit*. London: Goldsmiths, University of London.

Museums Association (2015). *Code of Ethics for Museums*. https://www.museumsassociation.org/ethics/code-of-ethics.

NEMO (Network of European Museum Organisations) (2020). *Final Report: Digitisation and IPR in European Museums*. https://www.ne-mo.org/news/article/nemo/nemo-report-on-digitisation-and-copyright-challenges-of-making-museum-collections-accessible-online.html.

Not, E.; Cavada, D.; Maule, S.; Pisetti, A.; Venturini, A. (2019). "Digital Augmentation of Historical Objects Through Tangible Interaction". *Journal on Computing and Cultural Heritage*, 12(3), Article 18, 19 pages. https://doi.org/10.1145/3297764.

Ozick, C. (1997). "Who Owns Anne Frank?". *The New Yorker.* https://www.newyorker.com/magazine/1997/10/06/who-owns-anne-frank.

The *Pierrot* (1911-15). The Museum of Childhood, Edinburgh. MC.86.86.

Pooley, S. (2015). "Children's Writing and the Popular Press in England 1876-1914". *History Workshop Journal*, 80(1), 75-98. https://doi.org/10.1093/hwj/dbv020.

Pratt, M.L. (1985). "Scratches on the Face of the Country; or, What Mr. Barrow Saw in the Land of the Bushmen". *Critical Inquiry*, 12, 119-43. https://doi.org/10.1086/448324.

Romein, C.A.; Kemman, M.; Birkholz, J.M. et al. (2020). "State of the Field: Digital History". *History*, 105, 291-312. https://doi.org/10.1111/1468-229X.12969.

Smithsonian Digitization Project Office. (2020). Smithsonian Digitization | 3D. https://3d.si.edu/.

Scollan, A.; Farini, F. (2020). "In, Out and Through Digital Worlds. Hybrid-transitions as a Space for Children's Agency". *International Journal of Early Years Education*, 28(1), 36-49. https://doi.org/10.1080/09669760.2019.1695586.

Shildrick, M. (1996). "Posthumanism and the Monstrous Body". *Body & Society*, 2(1), 1-15. https://doi.org/10.1177/1357034X96002001001.

Sloan, C. (2017). "'Periodicals of an Objectionable Character': Peers and Periodicals at Croydon Friends' School, 1826-1875". *Victorian Periodicals Review*, 50(4), 769-86. https://doi.org/10.1353/vpr.2017.0055.

Todorova, M. (2017). "Children's Voices from War Zones: Muted by Adult Mediation". *Bookbird: A Journal of International Children's Literature*, 55(2), 20-7.

Ventä-Olkkonen, L.; Iivari, N.; Kuutti, K. (2017). "Digital Technologies in Everyday Environments: Zooming in and Out to Children's and their Families' Smart Device Practices with Public and Private Screens". *Proceedings of the European Conference on Cognitive Ergonomics* (*ECCE 2017*). Association for Computing Machinery, New York, 129-36. https://doi-org.ezproxy.lib.gla.ac.uk/10.1145/3121283.3121302

Wolf, K.; Abdelhady, E.; Abdelrahman, Y.; Kubitza, T.; Schmidt, A. (2015). "MeSch: Tools for Interactive Exhibitions". *Proceedings of the Conference on Electronic Visualisation and the Arts (EVA'15)*, 261-9. https://doi.org/10.14236/ewic/eva2015.28.

Woodward, K. (1997). *Identity and Difference*. London: Sage.

Semestral journal

Venice Centre
for Digital and Public Humanities
Università Ca' Foscari Venezia

Università
Ca'Foscari
Venezia