The Alan Turing Institute

Humanities and data science special interest group

The challenges and prospects of the intersection of humanities and data science:

A White Paper from The Alan Turing Institute

Authors

Barbara McGillivray (The Alan Turing Institute, and University of Cambridge)
Beatrice Alex (University of Edinburgh)
Sarah Ames (National Library of Scotland)
Guyda Armstrong (University of Manchester)
David Beavan (The Alan Turing Institute)
Arianna Ciula (King's College London)
Giovanni Colavizza (University of Amsterdam)
James Cummings (Newcastle University)
David De Roure (University of Oxford)
Adam Farquhar
Simon Hengchen (University of Gothenburg)
Anouk Lang (University of Edinburgh)
James Loxley (University of Edinburgh)
Eirini Goudarouli (The National Archives, UK)
Federico Nanni (The Alan Turing Institute)
Andrea Nini (University of Manchester)
Julianne Nyhan (UCL)
Nicola Osborne (University of Edinburgh)
Thierry Poibeau (CNRS)
Mia Ridge (British Library)
Sonia Ranade (The National Archives, UK)
James Smithies (King's College London)
Melissa Terras (University of Edinburgh)
Andreas Vlachidis (UCL)
Pip Willcox (The National Archives, UK)

Citation information
McGillivray, Barbara et al. (2020). The challenges and prospects of the intersection of humanities and data science: A White Paper from The Alan Turing Institute.
Figshare. dx.doi.org/10.6084/m9.figshare.12732164

Contents

Summary

This paper was produced as part of the activities of the Humanities and Data Science Special Interest Group based at The Alan Turing Institute[1]. The group has created the opportunity for fruitful conversations in this area and has brought together voices from a range of different disciplinary backgrounds. This document shows an example of how conversations of this type can benefit and advance computational methods and understandings in and between the humanities and data science, bringing together a diverse community. We believe the Turing can act as a nexus of discussion on humanities and data science research at the national (and international) level, in areas such as education strategy, research best practices, and funding policy, and can promote and encourage research activities in this interdisciplinary area. Specific recommendations aimed at the Institute include:

– Allowing and encouraging PhD candidates from non-STEM backgrounds to be eligible to apply for the Turing enrichment scheme, thus enabling more collaborations at the intersection between humanities and data science;

– Identifying humanities as a priority area for the data science for Science programme[2] and include the phrase 'and Humanities' into the name of the programme;

– Joining existing training programmes aimed at digital humanities researchers and practitioners to provide data science skills, building on previous experience such as with the Digital Humanities at Oxford summer school[3].

– Ensuring representation and advocacy for the humanities in strategic and decision-making structures. This will stimulate diversity of engagements and impact across and promote further interdisciplinary work.

1 www.turing.ac.uk/research/interest-groups/humanities-and-data-science, (The Alan Turing Institute, 2020)
2 www.turing.ac.uk/research/research-programmes/data-science-science, (The Alan Turing Institute, 2020).
3 Researchers from the Turing convened two highly successful workshops in the 2019 edition of the Digital Humanities at Oxford Summer School (www.dhoxss.net), "From Text to Tech" and "Applied Data Analysis" and plans to continue such engagements are in place for future editions of the summer school.

Moreover, we outline the following more general recommendations to funders, academic institutions, and researchers to further support research at the intersection between humanities and data science:

---

1. Methodological frameworks and epistemic cultures.

We call for the use of a common methodological terminology in research at the intersection between humanities and data science, and for a wider use of shared research protocols across these domains. We recommend that authors make the methodological framework that they are using explicit in their publications, and we call for inclusive research practices to be fostered across research projects.

---

2. Best practices in the use and evaluation of computational tools.

We encourage practices that ensure transparency and openness in research, and training programmes that help to choose the most suitable computational tools and processes in humanities research. We also call for computational tools to be evaluated in a dialogue between data scientists and digital humanists.

---

3. Reproducible and open research.

We promote transparent and reproducible research in the humanities, covering data, code, workflows, computational environments, methods, and documentation. Research funders and academic institutions should put in place further incentives for humanities researchers to publish the digital resources, code, workflows and pipelines they create as legitimate research outputs, e.g. in the form of publications in data journals.

---

4. Technical infrastructure.

As data and computing requirements grow, a horizontal infrastructure should be developed in order to democratise access to digital resources and to guarantee their continued maintenance and improvement. We also recommend that institutions teaching and supporting digital humanities direct users to these shared infrastructures to promote their uptake.

5. Funding policy and research assessment.

We encourage the creation of cross-council schemes which fund collaborative data science projects, for example with humanities colleagues embedded in the teams from conception. In evaluation commissions, funding bodies should recognise interdisciplinary research as requiring to be evaluated by panels of experts themselves engaged in interdisciplinary research. Where appropriate, research protocols in data science projects concerning humanities data should allow for humanities perspectives (e.g. integration of data ethics issues and evaluation of machine learning results based on the needs of Humanities scholars). Institutions can invest in resources that bridge the gap between data scientists and digital humanities scholars, for example, by creating 'safe' spaces where practitioners across disciplines can create joint agendas for collaboration. Funders and research bodies supporting data science should ensure that their boards, and steering committees, comprise of those from a range of interdisciplinary backgrounds, including from the humanities, to encourage this dialogue to flourish.

6. Training, education, and expertise.

We acknowledge the need to upskill humanities researchers in quantitative and computational methods if they wish to, and to incorporate these methods in undergraduate and graduate degrees. We also recognise that people educated in the scientific disciplines would benefit from acquiring skills traditionally associated with the humanities. Consideration should be given to the development of robust talent pipelines, as well as short skills-enhancement courses and workshops, and university courses. Schemes for collaborative PhDs, internships and research secondments across disciplines, institutions, and businesses should be supported.

7. Career, development, and teams.

In a highly interdisciplinary research context, we encourage multiple career paths and working models so that students and early career researchers gain a sense of which career options might be open to them.

Preamble

The Humanities — academic disciplines that study aspects of human society and culture, such as history, linguistics, politics, divinity, and literatures — are faced with opportunities presented by digital tools and methods that could enable transformational innovative research. This document reflects on what data-driven research within the humanities entails and presents, priorities and recommendations for supporting and driving it forward. Our analysis applies particularly, although not exclusively, to the UK academic landscape.

This paper was produced as part of the activities of the Humanities and Data Science Special Interest Group based at the Turing[4] , the British national-level research institute for Data Science and Artificial Intelligence. Founded in 2017, the special interest group has grown to include over 80 people. Over three years of intensive collaborations and discussions (2017-2020), including community engagement events (such as the panel 'Data Science & Digital Humanities: new collaborations, new opportunities and new complexities' at the Digital Humanities 2019 conference[5]), the group has identified the need for a focused and coordinated reflection on a shared agenda for both the humanities and data science. This document is the result of those discussions, sustained and strengthened by a fruitful and inclusive exchange of expertise and viewpoints. While not all authors are directly affiliated with the Institute, the Turing provided the opportunity for these reflections to arise from an ongoing open conversation between researchers and practitioners.

The primary audience of this document consists of researchers and managers from the Institute, as well as researchers from different backgrounds (humanities, computer science, statistics, mathematics, social sciences), educators, policy makers, researchers and practitioners based in cultural heritage organisations (Galleries, Libraries, Archives and Museums - GLAM) and the creative sector. It will also be of interest to university leadership teams who may use it to inform longer-term strategic and planning decisions based on the potential of the humanities to play a part within cross-institution data science actions.

We have chosen the term 'Humanities and Data Science' in recognition of the diversity of activities, approaches and interdisciplinary collaborations encompassed by current work and discussion in this space. The more established term for computational approaches to humanities research, 'Digital Humanities', also appears throughout this document. While there is broad consensus around the scope of digital humanities as an area of research and teaching where epistemology is entangled with digital technology, there are multiple definitions[6]. Our definition of digital humanities is broad and inclusive, in alignment with Terras et al. (2013). For 'Data Science' we refer to the definition offered by the Turing[7] as the field that "brings together researchers in computer science, mathematics, statistics, machine learning, engineering and the social sciences" to study "the drive to turn [large amounts of] data into useful information, and to understand its powerful impact on science, society, the economy and our way of life". We recognise that there is a significant overlap between the fields of data science and artificial intelligence (AI). However, we have decided to leave the latter out of the scope of this paper, given its focus on the development of algorithms to perform actions in an autonomous way and the emphasis we want to place in the ability of data science methods to reveal new insights from humanities research data. We will also use the adjectives 'data-driven' and 'computational' to refer to certain aspects of the research process involving data science methods in the humanities.

4 www.turing.ac.uk/research/interest-groups/humanities-and-data-science, (The Alan Turing Institute, 2020).
5 https://doi.org/10.34894/B1UFVH
6 See whatisdigitalhumanities.com (Heppler, 2009-2014) for many differing viewpoints.
7 www.turing.ac.uk/about-us/frequently-asked-questions, (The Alan Turing Institute, 2020).

Why this paper, and why now

Since their beginnings, the digital humanities have engaged in an energetic debate about their scope, defining features, and relationship to the wider humanities, and have established themselves as a community of practice (Schreibman et al., 2004; Terras, 2010; Terras, 2013; Terras et al., 2013; Gold and Klein, 2016; The Digital Humanities Manifesto 2.0). The computational focus has characterised the field from its initial explorations (Hockey, 2004; Vanhoutte, 2013; Nyhan and Flinn, 2016) and the shift from the label 'Humanities Computing' to 'Digital Humanities' was a catalyst for change. In the history of the field, recurring cycles and productive tensions have arisen from the interfolding of computational methodologies and approaches with hermeneutic and critical modes of analysis (see McCarty, 2005; Rockwell and Sinclair, 2016; Jones, 2016). This document postulates that we are currently witnessing another one of these junctures, one that is calling for a critical involvement with data science.

In many ways, we are seeing earlier methods blending into, or being extended by data science. Digitisation workflows are being augmented with automatic information extraction, data analysis, automated transcription of handwritten documents[8], and visualisation of transcribed content[9]. Techniques developed for history, literary studies, and linguistics are being scaled towards larger datasets and more complex problems raising the bar of interpretability and questioning the validity of data collection and analysis methods. On the other hand, the field of data science has recently started to engage with non-STEM (Science, Technology, Engineering, and Mathematics) disciplines, by offering new data-driven modelling frameworks for addressing long-standing research questions (Kitchin, 2014; Lazer et al., 2009) and proposing so-called 'human-centred approaches' to data science, focussed on the interpretability of machine learning models and a more active role for human input in algorithms (See Chen et al., 2016).

Moreover, in the current historical context we are witnessing an increased awareness of the questions of diversity and inclusion in research and academia, and we are seeing the creation of a strong movement aimed at addressing such issues globally. We believe that this paper can play a role in reinforcing a positive message in this respect.

8 See for example transkribus.eu/Transkribus (Transkribus, 2020).
9 See for example the report on 'Post digitisation metadata enrichment' commissioned by Jisc (Digirati, 2019).

Context

Bridging the gap between humanities research and computational approaches has been one of the core aims of the humanities computing and digital humanities from their inception (See Nyhan and Passarotti, 2019, which explores one of a number of possible genealogies). Yet, over the past decade there has been an increased interest in developing and supporting activities that involve, as peer partners, communities of humanities scholars, the GLAM sector, creative computing and data science researchers, prompted in some cases by the availability of a great variety of large digital datasets for humanities research, coupled with emerging quantitative research frameworks and relatively cheap computing resources[10].

A number of initiatives, including targeted funding schemes (e.g. The Transatlantic Platform[11], Digging Into Data[12], and Digital Transformations[13]), have promoted efforts to answer humanities-related research questions, including methodological ones, through computational methods. Communities, groups, and committees have emerged to support these efforts over the years[14]. Examples include member organisations of the Alliance of Digital Humanities Organizations (ADHO)[15], the Methods Network[16], the Computational Humanities group in Leipzig[17], the Computational Humanities committee[18], Computational Humanities research[19], the AHRC-funded Computational Archival Science (CAS) research network[20], and the Advanced Information Collaboratory[21]. Such groups have also organised a number of events and workshops; a necessarily incomplete list includes ADHO-affiliated conferences from 1989 to the present day[22], Computational Humanities 2014[23], CAS: Exploring Data - Investigating Methodologies 2019[24], The Digital Experimentation Workshop series at The National Archives UK (2017 - today)[25], COMHUM 2018[26], and the Computational Humanities Research workshop[27]. Moreover, a series of publications (Schreibman et al. (eds), 2004; Siemens and Scheibman (eds), 2008; Hughes, 2008; Biemann et al., 2014; Ortolja-Baird et al., 2019; Jenset and McGillivray, 2017; among many others) have put forward epistemological reflections, methodological frameworks and approaches for conducting research in this area. Initiatives such as the HathiTrust Research Centre (2020)[28] and Stanford Literary Lab (2020)[29] are leading similar work in the USA. While there is no systematic analysis of the global landscape, initiatives such as Global Outlook: Digital Humanities (2020)[30] demonstrates the richness, heterogeneity and international breadth of the intersections between humanities research and computational approaches.

10 As already remarked for instance by Milligan (2012) when discussing a "third way of computational history".
11 The Transatlantic Platform is a collaboration between Humanities and social science research funders across South America, North America and Europe: www.transatlantic platform.com (2020).
12 The Digging into Data Challenge, a collaboration of international funders, aims to address how big data challenges change the landscape for social sciences and Humanities research: diggingintodata.org (The National Endowment for the Humanities, 2020).
13 The AHRC funded a number of projects under the Digital Transformations in Arts and Humanities theme intended explicitly to transform research methodologies in a number of areas.
14 Similar aims are shared by members of the TEI consortium (members.tei-c.org/Institutions) , or communities of domain-specialised practice such as Digital Medievalist (digitalmedievalist.wordpress.com) or Digital Classicist (www.digitalclassicist.org).
15 adho.org.
16 www.methodsnetwork.ac.uk/redist/pdf/finalreport.pdf.
17 ch.uni-leipzig.de/about.
18 www.ehumanities.nl/computational-humanities.
19 github.com/cohure/CoHuRe.
20 computationalarchives.net.
21 ai-collaboratory.net.
22 adho.org/conference.
23 www.dagstuhl.de/en/program/calendar/semhp/?semnr=14301.
24 blog.nationalarchives.gov.uk/computational-archival-science-cas-experimentation-knowledge-exchange-and-interdisciplinary-collaborations (Goudarouli, 2019).
25 www.nationalarchives.gov.uk/about/our-research-and-academic-collaboration/events-and-training/digital-experimentation-workshops (The National Archives, 2020).
26 wp.unil.ch/llist/en/event/comhum2018 (Laboratoire lausannois d'informatique et statistique textuelle, 2020)
27 cohure.github.io/CoHuRe (Arnold et al., 2020).
28 www.hathitrust.org/htrc.
29 litlab.stanford.edu.
30 www.globaloutlookdh.org.

From the OpenGlam movement encouraging open licensing of digitised content[31], to the open cultural data movement for museum APIs[32] to the collections as data movement encouraging libraries and archives to provide collections in machine-readable form[33], cultural heritage organisations have increasingly devoted attention to establishing infrastructures and services[34] that enable access and use of digital resources for humanities research, as well as employing computational and quantitative techniques to inform their own approaches. They are also increasingly embracing machine learning for processing mass digitised content[35] and analysis of collections as well as for process improvement[36] and innovative forms of discovery and interpretation[37].

A number of 'continuing education' programmes aim to fill the educational gap between the traditional offerings of museum studies, library and information studies and Humanities departments and the demands for computational and quantitative skills[38]. Examples of such programmes include:

1. General training events with multiple disparate workshops, such as the annual Digital Humanities training events at Oxford[39], Digital Humanities Summer Institute (DHSI) at Victoria[40], DHSI Atlantic at Cork[41], Humanities Intensive Learning and Teaching (HILT)[42], European Summer School in Digital Humanities in Leipzig[43];

2. Training events focussed on a specific topic or subject area, such as courses on Natural Language Processing methods and techniques[44], the Cambridge Cultural Heritage data school[45], the Qstep programme in Manchester, the new post-graduate certificate in Computing for Cultural Heritage[46], the Carpentries skills training series[47], the Helsinki Digital Humanities hackathon series[48];

3. Resources for online self-tuition such as Programming Historian[49], Digital Research Infrastructure for the Arts and Humanities (DARIAH) Teach and DARIAH Campus[50], Parthenos[51] Training Suite.

31 openglam.org
32 www.freshandnew.org/2010/10/launch-of-the-powerhouse-museum-collection-api-v1-at-amped and museum-id.com/unlocking-potential-next-open-cultural-data-museums-mia-ridge
33 collectionsasdata.github.io.
34 Theoretical and practical work for information integration in the field of Cultural Heritage has been undertaken for the past twenty years by initiatives such as the CIDOC Conceptual Reference Model (CRM, www.cidoc-crm.org) and Europeana (www.europeana.eu) with substantial impact in digital heritage.
35 For example: transkribus.eu/Transkribus.
36 For example: livingwithmachines.ac.uk, www.kb.nl/en/news/2019/kb-explores-artificial-intelligence-to-generate-metadata, themuseumsai.network and library.stanford.edu/projects/fantastic-futures.
37 For example British Library Labs (www.bl.uk/projects/british-library-labs), Library of Congress Labs (labs.loc.gov) and similar initiatives at the intersection with creative practitioners and industries.
38 In the UK The British Academy has devoted strategic focus to this; see www.thebritishacademy.ac.uk/tag/quantitative-skills-publications.
39 www.dhoxss.net.
40 dhsi.org.
41 www.ucc.ie/en/dhsiatlantic.
42 dhtraining.org/hilt/courses.
43 esu.culintec.de.
44 Teaching NLP for Digital Humanities, Teach4DH Workshop (ceur-ws.org/Vol-1918), the European Summer School in Logic, Language and Information (ESSLLI).
45 www.cdh.cam.ac.uk/dataschool/cultural-heritage-data-school.
46 www.bl.uk/projects/computingculturalheritage.
47 carpentries.org.
48 www.helsinki.fi/en/helsinki-centre-for-digital-Humanities/helsinki-digital-Humanities-hackathon, which is sometimes advertised via the DARIAH-CLARIN DH course registry (registries.clarin-dariah.eu/courses).
49 programminghistorian.org
50 teach.dariah.eu and campus.dariah.eu.
51 www.parthenos-project.eu.

On the other hand, there has been growing interest from computer scientists, physicists, and applied mathematicians to work on large data sets of Humanities data, such as text corpora, images, and others (Kitchin, 2014; Clifford et al., 2016). These datasets are interesting for researchers in these fields because of the complexity they often manifest. Humanities datasets are often unstructured, fragmentary, ambiguous, contradictory, multilingual, heterogeneous and bounded by the subjectivities of their data collection (e.g. Alex et al., 2016; Guiliano and Ridge, 2016) or limited by the data available or accessible to researchers at the time (e.g. Clifford et al., 2016; Hauswedell et al., 2020). Moreover, humanities datasets offer rich case studies for those interested in the statistical modelling of this messy and complex data (e.g. Underwood, 2018). Although this research has immense promise to yield new insights into the historical and cultural record, it is imperative that humanists and computer scientists engage in meaningful collaborations with each other in order to pursue it.

Both computational and humanistic domain knowledge are needed to engage with these datasets and their layers of complexity. Building on the strong tradition of interdisciplinary and intermural collaboration which has characterised the field of digital humanities, we need strong models of collaboration. Without such collaborations, there is a substantial risk that data-driven research does not say anything new or meaningful, repeats well-known distortions, or introduces new forms of bias at an even larger scale. In such a rich and vibrant landscape, this document aims to highlight a series of recommendations to help the communities involved interact in such a way as to reach the full potential of interdisciplinary research.

Scope

Research at the intersection between humanities and data science can take different shapes. Our conversation has been framed by the four general areas outlined below. In this document we focus more specifically on the first two, although some of our recommendations apply to all four and others not listed here:

– **Computational humanities research**. This area aims at applying existing methods and developing new methods to create and/or analyse digitised and born-digital datasets to answer both novel and established humanities research questions. Computational humanities research typically relies on quantifiable evidence and adopts computational and automatic procedures for processing and analysing data. Extensive work has been done to describe this area and we refer to external resources[52], as well as the lists of initiatives presented in the previous section for a fuller overview. Examples of this research have been undertaken on topics such as computational historical linguistics for the classical languages for example to find how the meaning of words changes over time and which factors can be identified for it (McGillivray et al., 2019), in the context of economic history cliometrics based on the Seshat database (Turchin et al., 2015), art analytics and modelling (Whitehouse et al., 2019; Fraiberger et al., 2018), the history of the Humanities (Colavizza, 2018), and web archival research (Nanni, 2018).

– **Infrastructure for Cultural Heritage**. This area is concerned with creating, storing and providing access to repositories of complex and nuanced digital (structured and unstructured) data from GLAM organisations for their use in research, as well as investigating the question of availability of digital resources as data, and accounting for biases and uncertainty in them. This infrastructure, whose solutions include closed commercial and openly available public systems, also enables the enrichment of cultural heritage data and metadata using state-of-the-art Data Science methods. Examples of research into existing and potential future infrastructure offerings include, the Living with Machines project[53], the University of Glasgow's GDD Network project, which explores the feasibility of creating a global register of digitised material[54], and the Collections as Data recommendations[55] (Padilla et al., 2019). Additional examples include the Big Data for Law project[56] and the new programme Towards a National Collection: Opening UK Heritage to the World[57], which brings together the Arts and Humanities Research Council (AHRC) with The Department for Digital, Culture, Media & Sport (DCMS) and AHRC Independent Research Organisations (IROs)[58] through the Strategic Priorities Fund led by UK Research and Innovation[59]. In addition, many national libraries also make collections data available on institutional websites, and/or offer API access to data[60]. There are efforts to develop robust, open-source tools enabling text mining and search of the mined output across large-scale text collections, e.g. the Defoe toolbox (Filgueira et al., 2019), to avoid humanities researchers having to develop ad hoc technological solutions.

52 For example: www.scottbot.net/HIAL/index.html@p=41533.html and cohure.github.io/CoHuRe .
53 www.livingwithmachines.ac.uk.
54 gddnetwork.arts.gla.ac.uk.
55 collectionsasdata.github.io.
56 Also see gtr.ukri.org/projects?ref=AH%2FL010232%2F1. This was previously known as the Legal Data Research Infrastructure and was funded by the UK AHRC under the 'digital transformations' call in 2012.
57 www.gov.uk/government/news/government-investment-backs-museums-of-the-future and ahrc.ukri.org/newsevents/news/first-awards-and-leadership-announced-towards-a-truly-national-collection/?utm_source=Twitter&utm_medium=social&utm_campaign=SocialSignIn&utm_content=.Social+Team%3A+funding+calls+and+events. This large funding effort has led to a number of projects. The National Archives, UK, received funding, through the Towards a National Collection programme, for two foundational projects: Deep Discoveries, a collaboration with the University of Surrey's Centre for Vision Speech and Signal Processing, V&A and Royal Botanic Garden Edinburgh, that aims to create a computer vision search platform that can identify and match images across digitised collections on a national scale; and, Engaging crowds: citizen research and heritage data at scale, a collaboration with the University of Oxford's Zooniverse, Royal Botanic Garden Edinburgh, and Royal Museums Greenwich, that aims at harnessing the capabilities of people-powered research to enrich understanding of cultural heritage collections through digitally enabled participation.
58 An IRO is an organisation which is deemed by AHRC to have a large enough research 'critical mass' to be considered for AHRC funding in the same way as a university. For more information, you can visit the IRO Consortium for the Arts and Humanities.
59 www.ukri.org/research/themes-and-programmes/strategic-priorities-fund.
60 For example, bl.iro.bl.uk, data.nls.uk, labs.kb.dk, lab.kb.nl, data.bnl.lu.

– **History and critique of data science**. This area analyses the characteristics of data science work, sometimes with a focus on the historicity of datasets, and tackles ethical and methodological questions aimed at improving current practices, for example, on issues such as diversity (D'Ignazio and Klein, 2020) and privacy. It also problematizes the very definition of data, considering their complexities, their inherent biases, their contextual and historical natures, in a critical and nuanced way (e.g. Drucker, 2011). Exemplary research in this area includes, among many others, MacKenzie (2017)'s study of the interface between machine learning and critical thought, Kaltenbrunner (2014; 2015)'s studies of infrastructure as a relational and emergent phenomenon that shapes data-driven humanities research and researchers, and the questions they can ask, in complex ways, and Noble (2018)'s study of how white patriarchy and algorithmic bias has resulted in the misrepresentation of women of colour and minorities in search engine results.

– **Algorithmic creativity and cultural innovation in the arts and humanities**. This area focuses on computational creativity, aiming to perform creative tasks with the aid of machines and to explore the plasticity of digital forms for delivering new radical ways of representation and mediation of the arts and humanities. This is the focus of various initiatives such as the Turing AI & Arts group[61] and the Creative Informatics programme in Edinburgh[62]. Individual artists are also increasingly including computational methods in their practice[63]. Examples of projects undertaken in collaboration with academic researchers, creative industries and the GLAM sector include the King's Digital Lab Digital Ghost Hunt[64] experience and the AI and Storytelling project[65].

61 www.turing.ac.uk/research/interest-groups/ai-arts.
62 creativeinformatics.org.
63 See, for example, www.forbes.com/sites/tabithagoldstaub/2018/09/24/machine-dreams-art-and-artificial-intelligence, thegradient.pub/the-past-present-and-future-of-ai-art, www.bl.uk/events/imaginary-cities, data.nls.uk/projects/artist-in-residence and www.mutualart.com/Article/The-Real-Future-of-Art-and-Artificial-In/D741A0C0C602F7E5.
64 digitalghosthunt.com.
65 www.kdl.kcl.ac.uk/our-work/ai-and-story-telling.

Challenges, opportunities, and recommendations

We envisage a future where humanities and data science researchers work together in synergy to realise the full potential of interdisciplinary work. But we recognise that there are cultural, practical, methodological, technical, financial, and infrastructural obstacles (Kemman, 2019) which currently slow down and in some cases stop this bidirectional exchange[66]. We therefore make the following series of recommendations, aimed at researchers in data science, humanities, GLAM practitioners and researchers, creative industries, funders, and policy makers. In some cases, when there are no definite answers, we recognise the challenges and aim to keep the conversation open while recommending that opportunities are created to reflect upon and address those challenges.

The recommendations below are presented into two groups: one related to the research process and one related to the support structures which are needed to enable it.

66 See also www.ukri.org/files/infrastructure/the-uks-research-and-innovation-infrastructure-opportunities-to-grow-our-capacity-final-low-res.

Research Process

1. Methodological frameworks and epistemic cultures

What does this mean? Incorporating new computational, quantitative, and data-driven approaches into the way humanities research is conducted requires us to articulate and rethink the whole research process in new ways (McCarty, 2005). Following a long tradition of methodological and theoretical work, we believe that dedicated efforts, tailored to the specific needs of the individual disciplines, would help keep this work current and relevant in today's fast-changing technological landscape. We welcome a plurality of voices into this discussion.

What is the issue/context? Some disciplines have started to pave the way in this direction: Jenset & McGillivray (2017) for historical linguistics, Graham et al. (2015) and McGillivray et al. (2018) for a preliminary account in history, Clifford et al. (2016) for environmental history, Bode (2018), Eve (2019) and Kuhn (2019) for literary studies and Smithies (2017) for the humanities more generally. However, more work is needed to develop general methodological reflections and research agendas, and to ensure these discussions and frameworks are fully inclusive (Ali, 2014; Earhart et al., 2016).

What do we propose and for whom? We call for a generous understanding of what counts as 'method', 'methodology', and 'methodological frameworks', alongside existing more explicit definitions. One practical way to achieve this is via a common terminology and a wider use of research protocols addressed jointly to both humanists and computational researchers. To facilitate this common understanding, we recommend that authors make the methodological framework used explicit in their publications, so that methodological decisions can be fully explained, motivated, and connected with the original research questions and interpretation of the research results. Aligned to this are issues of equality, diversity, and inclusion: we must ensure that a diverse range of voices (across the protected characteristics of race, religion or belief, age, gender reassignment, sex, sexual orientation, marriage and civil partnership and pregnancy and maternity, disability, and nation) are involved fully in humanities data science research, going forward, to avoid issues of digital colonialism, and avoid the perpetration of existing inequalities within the evolving research agenda (Risam, 2018).

2. Best practices in use and evaluation of computational tools

What does this mean? This concerns the practical workflows and the decisions that are made in the research process at the intersection between digital humanities and data science.

What is the issue/context? Using pre-existing tools to deploy computational techniques in humanities can result in work that is insufficiently critical of the quantitative methods on which such tools depend, see e.g. discussion in Nanni et al. (2016).

What do we propose and for whom? To address this challenge, we encourage the adoption of practices that ensure transparency (as well as reproducibility and openness, see next point) in the process and outputs of research, as well as the creation of training programmes that tackle statistical and computational literacy, and the question of suitability of computational tools and processes (including data quality, iterations and prototyping) in humanities research (see point below).

A related point concerns the development and evaluation of computational tools. In the same way as other computational fields have accepted that assumptions and generalisations need to be made in order to be able to evaluate methods and therefore foster new research, data-driven research in the humanities must be ready to critically examine assumptions and uncertainties in their evaluation benchmarks too (McGillivray et al., 2020). Moreover, it is important to always clarify when such tools are adopted for data exploration or for extracting quantitative evidence supporting an argument (Owens, 2012). In both settings, benchmarks should be built in a dialogue between data scientists and digital humanists in order to assess the methods' reliability and examine the types of errors to which they are prone. Eventually, when applicable, computational results should be made falsifiable to allow the community to assess their robustness and replicability (Nguyen et al., 2019; Tahmasebi and Hengchen, 2019). Engaging both humanists and data scientists with the same results is possible through methodologies such as "explicitly decoupling, for reviewing purposes, the assessment of computational results from those of interpretive work, to allow for broader engagement" (Colavizza, 2019). disability, and nation) are involved fully in humanities data science research, going forward, to avoid issues of digital colonialism, and avoid the perpetration of existing inequalities within the evolving research agenda (Risam, 2018).

3. Reproducible and open research

What does this mean? Work with computational methods occurs across a wide spectrum in the humanities, from rigorously empirical to experimental and exploratory or creative approaches (Smithies, 2017). Accordingly, different degrees of reproducibility — intended as the ability to reproduce comparable results with the same data and same analysis methods — apply to the differing research processes.

What is the issue/context? Reproducible and open research depends upon access to data and appropriate computational infrastructure. New computational methods and approaches are being undertaken in different computational environments, some of which are in closed and/or proprietary infrastructure and others in open and/or public infrastructure. Very often new and exciting computational methods are time-consuming and hard to implement, or have dependencies that are difficult to establish, raising additional barriers to making research reproducible. The desire for reproducibility follows a broader trend in scientific disciplines and data science in particular[67], and more investment is needed in regard to sustainability of code and open source frameworks. Journals such as the Journal of Open Humanities Data[68] and Research Data in the Humanities[69], which focus on the publication of digital research objects and their critical description, do exist, but they are niche venues and their awareness among humanities researchers is still low.

What do we propose and for whom? We promote transparent and reproducible research in the humanities, covering data, code, workflows (Liu, 2017), computational environments, methods and documentation. We encourage partnerships and initiatives involving humanities research groups and institutions like the Software Sustainability Institute[70] and the UK Research Software Engineer Association[71]. We encourage the uptake of existing open science principles, for example the FAIR Guiding Principles for scientific data management and stewardship, which ask that research data is Findable, Accessible, Interoperable, and Reusable (FAIR) (Wilkinson et al., 2016). Further incentives should be put in place for humanities researchers to publish the digital resources, datasets, data models, software modules, workflows and data pipelines they create as legitimate research outputs, for example in the form of publications in data journals.

67 www.turing.ac.uk/research/research-projects/turing-way-handbook-reproducible-data-science.
68 openhumanitiesdata.metajnl.com.
69 brill.com/view/journals/rdj/rdj-overview.xml.
70 www.software.ac.uk.
71 rse.ac.uk.

Enablers / support structures

4. Technical infrastructure

What is this about? This point concerns the large-scale digital infrastructure to host datasets and computational algorithms, to document and publish data models and workflows, and more in general to support research at the intersection between humanities and data science space.

What is the context? Extensive research policy efforts at the European level have been devoted to promoting investment in digital research infrastructures for the arts and humanities, and a number of large projects are active in this space. There has been a long-standing development of shared technological infrastructure in the digital humanities for over a decade (see DARIAH and CLARIN, and NEDIMAH for a methods taxonomy). However, there is a need to design and implement non-project dependent general, large-scale research infrastructures for the humanities (see Smithies, 2017) . This would give efficient access to developing technologies for humanities researchers, and encourage their uptake across humanities disciplines. This would also avoid the risk of fragmentation whereby individual projects build separate infrastructures that do not meet the requirements of FAIR, and are not maintainable and extensible over the long term. This need is shared by humanities researchers and by most cultural heritage organisations alike, since traditionally most humanities projects have been of a small-scale and low-resource nature.

What do we propose and for whom? As data and computing requirements grow, a horizontal infrastructure should be developed in order to democratise access to digital resources, and to guarantee their continued maintenance and improvement. We also recommend that institutions teaching and supporting digital humanities direct users to these shared infrastructures to promote their uptake. Some questions remain open: what are the requirements? which are general? and what are instead specific to the humanities? Which funding, governance, and implementation models should be pursued?

5. Funding policy and research assessment

What is this about? This point concerns the funding and research assessment landscape and how it can support research at the intersection between humanities and data science.

What is the context/issue? Humanities research provides a crucial lens for data science and has an essential role to play in addressing bias in data, inequalities, power structures, and social impact of data-driven work. With respect to research funding, Digging into Data and Trans-Atlantic Platform are examples of successful funding programmes. However, current budgetary constraints within single research councils still limit the support for collaborative research grants and hinder ambitious large-scale interdisciplinary projects which can bridge data science and the arts and humanities. In relation to research assessment, we welcome that digital research outputs can now be returned in the context of the Research Excellence Framework (REF) in the UK (Ciula, 2019). However, institutional recognition (by individual universities and departments, especially in relation to the REF) is also needed to legitimise forms of research outputs suited to collaborative computational work but do not have the same academic cachet as expected forms of publication in the Humanities, which is often the single-authored monograph .

What do we propose and for whom? Impactful interdisciplinary work needs specific support, and more recognition by funders, publishers, institutions and colleagues. This includes recognition that the time required for interdisciplinary collaboration can reduce the time available for publications and other traditional reward structures.

From the perspective of funding bodies, adequate cross-council schemes should be conceived which fund data science projects in genuinely collaborative teams, for example with humanities colleagues or hybrid roles embedded in the teams from conception. In evaluation commissions, funding bodies should recognise that interdisciplinary research requires its own evaluation by experts, and it should not be penalised by receiving an average of the scores given by reviewers from different fields (See Ranjbaran and Marras, 2011).

From the point of view of researchers, protocols should be developed for data science projects which allow for the intersection with humanities perspectives such as integration of data ethics issues and evaluation of machine learning results relying on criteria established in collaboration with humanities scholars. We recognise the importance of institutional investment into resources that bridge the gap between data scientists and digital humanities scholars. We can achieve this through creating ‘safe’ spaces where practitioners across disciplines can establish a common language, develop mutual interests, and create joint agendas for collaboration. This is fundamental to establish a coherent humanities and data science vision, and to influence funding bodies’ policies to provide opportunities for interdisciplinary exploration and experimentation. Finally, we encourage funding councils to embrace interdisciplinarity in their own business: encouraging cross-fertilisation from the arts, humanities, social sciences, and sciences, at board and steering group level, allowing meaningful discussion that can embed interdisciplinary drivers into institutional practices.

6. Training, education, and expertise

What is this about? The growing availability of digital and digitised data relevant to humanists is increasingly allowing scholars to engage with both old and new questions using quantitative and computational methods. Furthermore, societal challenges calling for humanistic inquiry are often tied with the development of novel digital technologies. As technological approaches change and develop, it is crucial that the digital humanities community keep up to date with new skills, both to engage fully with their affordances, but also to experiment with how they can be applied to the humanities. In parallel, skills traditionally associated with the humanities are increasingly in demand among data scientists and are still lacking from most of the scientific education programmes.

What is the context? Traditional humanities education is often based on close, interpretative reading of dense texts and other artefacts, and also based on a high degree of specialisation. Consequently, while the humanities have developed a core set of methods and techniques for the rigorous interpretation of their sources, traditionally they lack training in the core subjects of modern data, computer and information science: mathematics, statistics and computer programming. To avoid becoming users of black-box tools and techniques, humanities scholars, teachers and professionals need to broaden their training and expertise by integrating or expanding it beyond the boundaries of the educational offering of their disciplines. This need is not fundamentally different from that of other academic disciplines, such as medicine, life sciences and the social sciences. The demand for computational and quantitative skills among digital humanities scholars, GLAM professionals and humanities students can be seen at different levels. Some humanists wish to learn about the basic terminology and workings of quantitative and computational concepts but would prefer to collaborate with computer scientists rather than learn to program or do statistical analyses themselves (See Cummings, 2019). Other humanists are actively interested in complementing their skillset with data, computer, and information science. On the other hand, there is an increasing need for data scientists with non-humanities backgrounds to question some of their assumptions about data. This comes with the understanding that 'humanities questions' can help them do research which grapples with the complex reality of human society and culture in deeper and more useful ways. Moreover, we recognise the need to train data scientists in understanding the added value of humanities research, so that those complex problems can be tackled in their complexity. One example is the recent research on detecting word meaning change in language. This has been tackled by computational linguists effectively, but by making strong simplifying assumptions, which means that humanists do not engage with this research.

What do we propose and for whom? The challenge here is to find a way to train and upskill humanities researchers in quantitative and computational methods, while at the same time incorporating the basic principles from these methods throughout undergraduate and graduate degrees, so humanities graduates are well equipped to lead projects but also potentially undertake careers in research software engineering and data science for arts and humanities. Consideration should be given to the development of robust talent pipelines, as well as short skills-enhancement courses and workshops, and university courses. A set of basic courses in data science and software engineering, ideally shared across the community, while not turning into a full programme, would offer the foundational skills to support humanists in having structured and informed conversations with computer scientists and data scientists needed in interdisciplinary projects. In addition, schemes to sustain collaborative PhDs, internships and research secondments across disciplines, institutions and businesses should be supported. At the same time, we need to open channels of communication between computational research in the humanities and 'mainstream' humanities fields and promote cross-disciplinary discussions. While a subsection of researchers develops new tools and methods at the leading edge of technology, we need more established digital methods to be accepted, evaluated, and incorporated by humanities disciplines, given the benefits they can offer to established lines of humanities inquiry. Theorised approaches to these digital methods will also help inform evaluation criteria for new tools and methods (see Best Practices section).

7. Career, development, and teams

What is this about? When basic data science principles and quantitative methods are woven into undergraduate and postgraduate humanities curriculums, we will start to see cohorts of people with the 'blended' scholarly and technical profiles necessary for advanced methods, which will bring new perspectives to industry positions as data scientists and data analysis. They will also bring technical skills to the careers traditionally chosen by humanities graduates, such as civil service, management, human resources, but also marketing, publishing, education, and so on. If we focus on careers in universities and cultural institutions, these people will be capable of undertaking advanced degrees in data science and the humanities, and would ideally be able to choose two pathways: a scholarly route into a traditional academic career, or a more technical route into a professional career in data science or Research Software Engineering (RSE) . Those who choose the latter route will be in high demand and will join teams who are beginning to be offered quality career paths (aligned to both academia and industry) and designing team structures tailored to data science and the humanities. Both routes will offer career opportunities outside academia, across all sectors (Health, IT, Government, Finance, NGO etc) and organisational functions (Operations, Management, Policy Analysis etc.), ensuring the humanities retain their position as a major contributor to all aspects of civil and commercial society and culture.

What is the context/issue? RSE career paths and models for RSE teams, despite being central to UK e-infrastructure strategy , are nascent. On the other hand, digital humanities curricula offer very different models across Europe and globally. While this diversity is important for the field to expand and thrive, to sustain career pipelines and profiles suitable to work at the crossroads between humanities and data science the next generation needs training programmes geared to practice and process-oriented methodologies . The current generation of RSEs working at the intersection of data science and the humanities need to help define their own career pathways and ways of working, but also foster new generations. A combined, and collaborative, approach to both education and career development is needed at the project, institutional, national, and international level. At the same time, from the point of view of academic digital humanists, the time and investment needed to acquire technical skills can often mean that digital humanities CVs and careers are atypical from traditional humanities arcs, and the necessary continual upskilling in the digital domain is often orthogonal to a traditional academic humanities career of increasing specialism, which results in issues regarding employment, promotion, and tenure.

What do we propose and for whom? We need to encourage multiple career paths and working models, so that students gain a sense of which career options are open to them, and the current generation of RSEs (and their managers) have resources they can adapt to the local circumstances. Examples exist (See Smithies, 2019), but more needs to be done to encourage teams and their institutions to think critically about the RSE role, problematise the issues, and ensure they are aligned to workforce needs in Higher Education (HE) and the creative, government, and commercial sectors: How do we implement financial models and incentive schemes that encourage the kind of flexible working RSEs need? How do we make an RSE career in data science and the humanities appealing enough to attract people who might otherwise choose to apply their skills to STEM disciplines, or work outside the research community? How do we increase diversity in RSE, so creativity, research quality, and innovation improve? How do we balance the need for rigorous and transparent engineering processes with the need for flexibility, creativity, and acceptance of failure?

Conclusion

In this document we have outlined a series of recommendations which will hopefully support research at the intersection between data science and humanities and allow it to advance in new ways. We also welcome the support we have received by the Institute in exploring the recommendations presented here. We see this document as a marker in the sand of what has been achieved so far and we acknowledge the extraordinary potential for ground-breaking new research at the intersection between data science and humanities. This research has intangible value and outputs that can contribute to the development of the UK culture and industry, strengthening the value of intellectual work and collaborations in the service of shared cultural activities.

We would like to end this document with a note on the nature of this interdisciplinary exchange. We believe that this interdisciplinary exchange is bidirectional. In fact, we observe that much of the debate around engaging the humanities with possibilities in data science happens with what we could call a 'techno-futurist' approach: bringing the promise of data science to the humanities. However, humanities researchers have their own skill sets, value systems, methods, expertise and approaches which the data science community can learn from, and this should be a two-way exchange of approaches and knowledge.

Humanists combine rich knowledge of their own field and expertise in analysing and interpreting gaps in the data, including the significance of negative results. Training in the humanities provides skills which allows researchers to excel in the detection and confrontation of bias, the analysis of tacit power structures, ethical, feminist, and non-capitalist approaches to information flow and data analysis, and close-reading approaches to placing the human, and human society, at the centre of debates. These are the skills which data science is - latterly - beginning to acknowledge that the industry and research needs, in order to build ethical, supportive technologies that can be classed as 'data science for social good' (although that is rife with its own value judgements).

Building on its position as the UK national institute for Data Science and Artificial Intelligence, we propose that the Turing acts as a nexus of the discussion on humanities research at the national (and international) level, engaging on topics such as education strategy, research best practices, and funding policy. The recommendations contained in this document are a first step in this direction, and we aim to develop them further through specific activities and more publications in the future.

Acknowledgements

We would like to thank Bethany Johnstone for her help in copyediting this text.

This work was supported by The Alan Turing Institute under the EPSRC grant EP/N510129/1.

References

Alex, B., Grover, C., Oberlander, J., Thomson, T., Anderson, M., Loxley, J., Hinrichs, U. and Ke Zhou, Ke. (2016). Palimpsest: Improving assisted curation of loco-specific literature. In: Digital Scholarship in the Humanities, 32(1), pp.i4–i16.

Alex, B., Alexander, A., Beavan, D., Goudarouli, E., Impett, L., McGillivray, B., McGregor, N., & Ridge, M. (2019). Data Science & Digital Humanities: New collaborations, new opportunities and new complexities. Digital Humanities 2019, Utrecht, The Netherlands. staticweb.hum.uu.nl/dh2019/dh2019.adho.org/panels/index.html

Ali, M (2014). Towards a decolonial computing. In: Ambiguous Technologies: Philosophical Issues, Practical Solutions, Human Nature, International Society of Ethics and Information Technology. pp. 28-35.

Biemann, C., Crane, G., Fellbaum, C. and Mehler, A. (eds) (2014). Computational Humanities - bridging the gap between Computer Science and Digital Humanities. Dagstuhl Seminar 14301. http://cpb-us-w2.wpmucdn.com/u.osu.edu/dist/4/27964/files/2016/01/DagstuhlSeminarFinalReport-2a7n3h7.pdf

Bode, K. (2018). A World of Fiction: Digital Collections and the Future of Literary History. USA: The University of Michigan Press. https://doi.org/10.3998/mpub.8784777

Chen, N., Kocielnik, R., Drouhard, M., Peña-Araya, V., Suh, J., Cen, K., Zheng, X. and Aragon, C. R. (2016). Challenges of Applying Machine Learning to Qualitative Coding. In: CHI 2016 workshop on Human Centred Machine Learning (HCML 2016)

Ciula, A. (2019). KDL Checklist for Digital Outputs Assessment (Version 2.0). Zenodo. http://doi.org/10.5281/zenodo.3361580

Clifford, J., Alex, B., Coates, C., Watson, A. and Klein, E. (2010). Geoparsing History: Locating Commodities in Ten Million Pages of Nineteenth-Century Sources. In: Historical Methods, 49(3), pp.115–131. http://doi.org/10.1080/01615440.2015.1116419

Colavizza, G. (2019). Are we breaking the social contract? In: Journal of Cultural Analytics. September, pp.1-10. http://doi.org/10.22148/001c.11828

Colavizza, G. (2018). Understanding the History of the Humanities from a Bibliometric Perspective: Expansion, Conjunctures, and Traditions in the Last Decades of Venetian Historiography (1950–2013). In: History of Humanities, 3(2), pp.377–406. https://doi.org/10.1086/699300

Cummings, J. (2019) Opening the book: data models and distractions in digital scholarly editing. In: International Journal of Digital Humanities, 1, pp.179–193. https://doi.org/10.1007/s42803-019-00016-6

D'Ignazio, C., Klein, L. (2020). Data Feminism. USA: MIT Press

Digirati. (2019). Post digitisation metadata enrichment tools evaluation report. Jisc. https://digitisation.jiscinvolve.org/wp/files/2019/12/SM20C-Evaluation-Report-v-1.1-for-distribution.pdf

Drucker, J. (2011). Humanities Approaches to Graphical Display. In: Digital Humanities Quarterly, 5(1), pp.1-52. www.digitalhumanities.org/dhq/vol/5/1/000091/000091.html

Earhart, A., Gil, A., Risam, R., Bordalejo, B., Galina, I., Hughes, L., Terras, M. (2016). Quality Matters: Diversity and the Digital Humanities. In Digital Humanities 2016: Conference Abstracts. Kraków: Jagiellonian University & Pedagogical University, pp. 61-63. http://dh2016.adho.org/abstracts/14

Eve, M. P. (2019). Close Reading with Computers: Textual Scholarship, Computational Formalism, and David Mitchell's Cloud Atlas. Stanford: Stanford University Press.

Filgueira, R., Jackson, M., Terras, M., Beavan, D., Roubickov, A., Hobson, T., Ardanuy, M.C., Colavizza, G., Krause, A., Hetherington, J., Hauswedell, T. (2019). defoe: A Spark-based Toolbox for Analysing Digital Historical Textual Data. In: 2019 IEEE 15th International Conference on e-Science (e-Science). IEEE 15th International Conference on e-Science (e-Science). USA: San Diego

Fraiberger, Samuel P., Roberta Sinatra, Magnus Resch, Christoph Riedl, and Albert-László Barabási. (2018). Quantifying Reputation and Success in Art. In: Science, 362(6416), pp.825-829. https://doi.org/10.1126/science.aau7224

Gold, M. K. and Klein, L. F. (2016). Debates in the Digital Humanities. USA: University of Minnesota Press.

Goudarouli, E., (accepted, to be published 2020). From Hilary Jenkinson to Artificial Intelligence Era: How digital innovation reshapes archival thinking and practice. A. Wiggins and A. Prescott (eds.) In: Archives: Power, Truth and Fiction, the 21st Century Approaches to Literature. Oxford: Oxford University Press.

Goudarouli, E., Sexton, A., Sheridan, J. (2018). The Challenge of the Digital and the Future Archive: Through the Lens of The National Archives UK. In: Philosophy and Technology, 32(1), pp.173-183. https://doi.org/10.1007/s13347-018-0333-3

Graham, Shawn, Milligan, Ian and Weingart, Scott (2015). Exploring Big Historical Data: The Historian's Macroscope. Imperial College Press. https://doi.org/10.1142/p981

Grosz, B.J., Grant, D.G., Vredenburgh, K., Behrends, J., Hu, L., Simmons, A. and Waldo, J., (2019). Embedded EthiCS: integrating ethics across CS education. In: Communications of the ACM, 62(8), pp.54-61.

Guiliano, J., & Ridge, M. (2016). The Future of Digital Methods for Complex Datasets: An Introduction. In: International Journal of Humanities and Arts Computing, 10(1), pp.1–7. https://doi.org/10.3366/ijhac.2016.0155

Hockey, S. (2004). The History of Humanities Computing. In: Schreibman, S., Siemens, R., Unsworth, J. (eds), A Companion to Digital Humanities. Oxford: Blackwell. www.digitalhumanities.org/companion/view?docId=blackwell/9781405103213/9781405103213.xml&chunk.id=ss1-2-1

Hughes, L. (ed.) (2008). The AHRC ICT Methods Network. London: Office for Humanities Communication. www.methodsnetwork.ac.uk/redist/pdf/finalreport.pdf

Jones, S. E., (2016). Busa, R. S. J. and the Emergence of Humanities Computing: The Priest and the Punched Cards. London: Routledge. www.routledge.com/products/9781138186774

Kaltenbrunner, W. (2014). Infrastructural Inversion as a Generative Resource in Digital Scholarship. In: Science as Culture, 0(0), pp.1–23. https://doi.org/10.1080/09505431.2014.917621

Kaltenbrunner, W. (2015). Scholarly Labour and Digital Collaboration in Literary Studies. In: Social Epistemology, 29(2), pp.207–33. https://doi.org/10.1080/02691728.2014.907834

Kemman, M. (2019) Boundary Practices of Digital Humanities Collaborations. In: DH Benelux journal, 1, pp.1-24.

Kitchin, R. (2014). Big Data, new epistemologies and paradigm shifts. In: Big data & society, 1(1), pp.1-12. https://doi.org/10.1177/2053951714528481

Kuhn, J. (2019). Computational text analysis within the Humanities: How to combine working practices from the contributing fields? In: Language Resources & Evaluation, 53, pp.565–602. https://doi.org/10.1007/s10579-019-09459-3

Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabási, A. L., Brewer, D., ... & Jebara, T. (2009). Computational social science. In: Science, 323(5915), pp.721-723.

Liu, A. (2017). Assessing Data Workflows for Common Data 'Moves' Across Disciplines. May 2017. http://doi.org/10.21972/G21593

Mackenzie, A (2017). Machine Learners: Archaeology of a Data Practice. USA: The MIT Press.

McCarty, W. (2005). Humanities Computing. UK: Palgrave Macmillan.

McGillivray, B. (2014). Methods in Latin Computational Linguistics. Leiden: Brill.

McGillivray, B., Colavizza, G., and Blanke, T. (2018). Towards a Quantitative Research Framework for Historical Disciplines. In: Piotrowski, M. (ed), COMHUM 2018: Book of Abstracts for the Workshop on Computational Methods in the Humanities 2018, Université de Lausanne, June 4-5, pp.29-31. Zenado. https://zenodo.org/record/1312779#.W2B4I62ZNTY

McGillivray, B., Hengchen, S., Lähteenoja, Palma, M., Vatri, A. (2019). A computational approach to lexical polysemy in Ancient Greek, In: Digital Scholarship in the Humanities, 34(4), pp.893-907. UK: Oxford University Press. https://doi.org/10.1093/llc/fqz036

McGillivray, B., Poibeau, T. and Ruiz Fabo, P. (2020). Digital Humanities and Natural Language Processing: Je t'aime...moi non plus. In: Digital Humanities Quarterly, 14(2), pp.1-45. www.digitalhumanities.org/dhq/vol/14/2/000454/000454.html

McGillivray, B. (accepted, to be published 2020). Computational methods for semantic analysis of historical texts. In: Schuster, K. and Dunn, S. (eds). Routledge International Handbook of Research Methods in Digital Humanities. London: Routledge.

Milligan, I. (2012). Mining the 'Internet Graveyard': rethinking the historians' toolkit. In: Journal of the Canadian Historical Association/Revue de la Société historique du Canada, 23(2), pp.21-64. https://doi.org/10.7202/1015788ar

Nanni, F., Kümper, H., and Ponzetto, S. P. (2016). Semi-supervised textual analysis and historical research helping each other: Some thoughts and observations. In: International

Nanni, F., (2018). Collecting primary sources from web archives: A tale of scarcity and abundance. In: Brugger, N. and Milligan, I. (eds). The SAGE Handbook of Web History. UK: SAGE publications Limited.

Nguyen, D., Liakata, M., DeDeo, S., Eisenstein, J., Mimno, D., Tromble, R. and Winters, J., (2019). How we do things with words: Analysing text as social and cultural data. arXiv preprint arXiv:1907.01468. pp.1-24. https://arxiv.org/abs/1907.01468

Noble, S. U. (2018). Algorithms of Oppression: How Search Engines Reinforce Racism. New York: NYU Press.

Nyhan, J. and Passarotti, M. (2019). One origin of Digital Humanities: Fr Roberto Busa in his own words. Switzerland: Springer Nature.

Nyhan, J. and Flinn, A. (2016). Computation and the Humanities: Towards an Oral History of Digital Humanities. Switzerland: Springer.

Ortolja-Baird, A., Pickering, V., Nyhan, J., Sloan, K. and Fleming, M. (2019). Digital Humanities in the Memory Institution: The Challenges of Encoding Sir Hans Sloane's Early Modern Catalogues of His Collections. In: Open Library of Humanities, 5(1), pp.44. http://doi.org/10.16995/olh.409

Owens, T. (2012). Discovery and justification are different: Notes on science-ing the humanities. Personal Website: www.trevorowens.org/2012/11/discovery-and-justificationare-different-notes-on-sciencing-the-humanities

Padilla, T., Allen, L., Frost, H., Potvin, S., Russey R. E., and Varner, S. (2019). Final Report --- Always Already Computational: Collections as Data (Version 1). Zenodo. http://doi.org/10.5281/zenodo.3152935

Piotrowski, M. (2012). Natural Language Processing for Historical Texts. Morgan & Claypool Publishers. In: Synthesis lectures on human language technologies, 5(2), pp.1-157. https://doi.org/10.2200/S00436ED1V01Y201207HLT017

Piotrowski, M. (2018). Digital Humanities: An Explication. In: Burghardt, M. & Müller-Birn, C. (eds.), INF-DH-2018. Bonn: Gesellschaft für Informatik e.V. https://dx.doi.org/10.18420/infdh2018-07

Prisoner, T. (n.d), The Digital Humanities Manifesto 2.0. In: Humanities Blast Publications. www.humanitiesblast.com/manifesto/Manifesto_V2.pdf

Ranjbaran, Farzam and Marras, Cristina (2011) European Peer Review Guide: Integrating Policies and Practices into Coherent Procedures. ESF Member Organisation Forum on Evaluation of Publicly Funded Research. Strasbourg. http://repository.fteval.at/148

Risam, R. (2018). New Digital Worlds Postcolonial Digital Humanities in Theory, Praxis, and Pedagogy. USA: Northwestern University Press.

Rockwell, G. and Sinclaire, S. (2016). Hermeneutica: Computer-Assisted Interpretation in the Humanities. USA: The MIT Press,.

Schreibman, S., Siemens, R. and Unsworth, J. (2004). A Companion to Digital Humanities. Oxford: Blackwell. www.digitalhumanities.org/companion

Smithies, J. (2017) Software Intensive Humanities. In The Digital Humanities and the Digital Modern. Basingstoke: Palgrave Macmillan.

Smithies, J. (2017) Towards a Systems Analysis of the Humanities. In: The Digital Humanities and the Digital Modern. Basingstoke: Palgrave Macmillan.

Smithies, J. (2019). Research Software (RS) Careers: Generic Learnings from King's Digital Lab, King's College London. Zenodo. https://doi.org/10.5281/zenodo.2564790

Smithies, J. (2017). The Digital Humanities and the Digital Modern. Basingstoke: Palgrave Macmillan.

Tahmasebi, N. and Hengchen, S. (2019). The Strengths and Pitfalls of Large-Scale Text Mining for Literary Studies. In: Samlaren: tidskrift för svensk litteraturvetenskaplig forskning, 140, pp.198-227. www.diva-portal.org/smash/get/diva2:1415010/FULLTEXT01.pdf

Terras, M., Nyhan, J. and Vanhoutte, E. (2013). Defining Digital Humanities: A Reader. London: Routledge.

Terras, M. (2010). The Digital Classicist: Disciplinary Focus and Interdisciplinary Vision. In: Bodard, G., Mahoney, S. (Eds.). Digital Research in the Study of the Classical Antiquity. pp. 171-191. London: Routledge.

Terras, M. (2012). Being the Other: Interdisciplinary work in Computational Science and the Humanities. In: McCarty, W., Deegan, W. (Eds.), Collaborative Research in the Digital Humanities. pp. 213-230. London: Ashgate.

Turchin, P., Brennan, R., Currie, T., Feeney, K., Francois, P., Hoyer, D. and Manning J., Marciniak, A., Mullins, D., Palmisano, A., Peregrine, P., Turner, E. and Whitehouse, H. (2015). Seshat: The Global History Databank. In: Cliodynamics: The Journal of Quantitative History and Cultural Evolution, 6(1). https://doi.org/10.21237/C7CLIO6127917

Underwood, T. (2018). Why an Age of Machine Learning Needs the Humanities. Public Books [blog]. December 5, 2018. www.publicbooks.org/why-an-age-of-machine-learning-needs-the-humanities

Vanhoutte, E. (2013). The Gates of Hell: History and Definition of Digital | Humanities | Computing. In Terras, M., Nyhan J. & Vanhoutte, E. (eds.) Defining Digital Humanities: A Reader. Farnham: Ashgate Publishing, pp.119-156.

Whitehouse, H., François, P., Savage, P., Currie, T., Feeney, K., Cioni, E., Purcell, P., Ross, R., Larson, J., Baines, J., Haar B., Covey, A. and Turchin, P. (2019). Complex Societies Precede Moralizing Gods throughout World History. In: Nature, 568, pp.226-229. https://doi.org/10.1038/s41586-019-1043-4

Wilkinson, Mark, Dumontier, Michel, Aalbersberg, Ijsbrand et al. The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data 3, 160018 (2016). https://doi.org/10.1038/sdata.2016.18

turing.ac.uk
@turinginst