ADVANCING DIGITAL HUMANITIES

Research, Methods, Theory

Edited by **Paul Longley Arthur**
and **Katherine Bode**

Advancing Digital Humanities

This page intentionally left blank

Advancing Digital Humanities

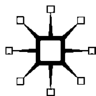
Research, Methods, Theories

Edited by

Paul Longley Arthur
University of Western Sydney, Australia

Katherine Bode
Australian National University, Australia

palgrave
macmillan



Introduction, selection and editorial matter © Paul Longley Arthur and Katherine Bode 2014
Individual chapters © Respective authors 2014
Softcover reprint of the hardcover 1st edition 2014 978-1-137-33699-6

All rights reserved. No reproduction, copy or transmission of this publication may be made without written permission.

No portion of this publication may be reproduced, copied or transmitted save with written permission or in accordance with the provisions of the Copyright, Designs and Patents Act 1988, or under the terms of any licence permitting limited copying issued by the Copyright Licensing Agency, Saffron House, 6–10 Kirby Street, London EC1N 8TS.

Any person who does any unauthorized act in relation to this publication may be liable to criminal prosecution and civil claims for damages.

The authors have asserted their rights to be identified as the authors of this work in accordance with the Copyright, Designs and Patents Act 1988.

First published 2014 by
PALGRAVE MACMILLAN

Palgrave Macmillan in the UK is an imprint of Macmillan Publishers Limited, registered in England, company number 785998, of Houndmills, Basingstoke, Hampshire RG21 6XS.

Palgrave Macmillan in the US is a division of St Martin's Press LLC, 175 Fifth Avenue, New York, NY 10010.

Palgrave Macmillan is the global academic imprint of the above companies and has companies and representatives throughout the world.

Palgrave® and Macmillan® are registered trademarks in the United States, the United Kingdom, Europe and other countries.

ISBN 978–1–137–33700–9 paperback
ISBN 978-1-137-33700-9 ISBN 978-1-137-33701-6 (eBook)
DOI 10.1057/9781137337016
This book is printed on paper suitable for recycling and made from fully managed and sustained forest sources. Logging, pulping and manufacturing processes are expected to conform to the environmental regulations of the country of origin.

A catalogue record for this book is available from the British Library.

Library of Congress Cataloging-in-Publication Data
Advancing digital humanities : research, methods, theories / edited by Katherine Bode (Australian National University, Australia), Paul Longley Arthur (University of Western Sydney, Australia).
pages cm
Summary: "Advancing Digital Humanities moves beyond definition of this dynamic and fast growing field to show how its arguments, analyses, findings and theories are pioneering new directions in the humanities globally. Sections cover digital methods, critical curation and research futures, with theoretical and practical chapters framed around key areas of activity including modelling collections, data-driven analysis, and thinking through building. These are linked through the concept of 'ambitious generosity,' a way of working to pursue large-scale research questions while supporting and enabling other research areas and approaches, both within and beyond the academy" — Provided by publisher.
Includes bibliographical references.
ISBN 978-1-137-33700-9
ISBN 978–1–137–33700–9 (paperback)
1. Humanities—Data processing. 2. Humanities—Methodology. 3. Humanities—Research. 4. Digital media. 5. Digital communications. 6. Information storage and retrieval systems—Humanities. I. Bode, Katherine. II. Arthur, Paul Longley.
AZ105.A35 2014
001.30285—dc23 2014024813

Contents

List of Figures and Tables vii

Notes on Contributors x

1 Collecting Ourselves 1
Katherine Bode and Paul Longley Arthur

Part I Transforming Disciplines

2 Exercises in Battology: Digitizing Samuel Beckett's *Watt* 15
Mark Byron

3 Stylometry of Dickens's Language: An Experiment with Random Forests 28
Tomoji Tabata

4 Patterns and Trends in Harlequin Category Romances 54
Jack Elliott

5 The Printers' Web: New Tools to Crack Old Chestnuts 68
Sydney J. Shep

6 Biographical Dictionaries in the Digital Era 83
Paul Longley Arthur

Part II Media Methods

7 Digital Methods in New Cinema History 95
Richard Maltby, Dylan Walker, and Mike Walsh

8 A 'Big Data' Approach to Mapping the Australian Twittersphere 113
Axel Bruns, Jean Burgess, and Tim Highfield

9 iResearch: What Do Smartphones Tell Us about the Digital Human? 130
Mark Coté

10 Screenshots as Virtual Photography: Cybernetics, Remediation, and Affect 141
Christopher Moore

Part III Critical Curation

11 Rethinking Collections 163
Julia Flanders

12 Methods and Canons: An Interdisciplinary Excursion 175
Katherine Bode and Tara Murphy

13 Reading the Text, Walking the Terrain, Following the Map: Do We See the Same Landscape? 194
Øyvind Eide

14 Doing the Sheep Good: Facilitating Engagement in Digital Humanities and Creative Arts Research 206
Deb Verhoeven

15 Materialities of Software: Logistics, Labour, Infrastructure 221
Ned Rossiter

Part IV Research Futures

16 Digital Humanities: Is Bigger, Better? 243
Peter Robinson

17 Margins, Mainstreams and the Mission of Digital Humanities 258
Paul Turnbull

18 The Big Bang of Online Reading 274
Alan Liu

19 Getting There from Here: Remembering the Future of Digital Humanities 291
Willard McCarty

Index 322

Figures and Tables

Figures

2.1 Juxta window of *Watt* MS 6:98–100 and *Watt* (1953/1959, 238), with collation histogram and comment apparatus 22
3.1 A multidimensional scaling plot based on the proximity matrix generated by RF: Dickens versus Collins 31
3.2 A multidimensional scaling plot based on the proximity matrix generated by RF: Dickens versus reference corpus 33
3.3 Result of a COHA search: *with his/her eyes fixed on/upon* 37
3.4 Result of a COHA search: *with his/her eyes on/upon* 37
3.5 Result of a COHA search: *with his hands in his* 39
3.6 Result of a COHA search: *his hands in his pockets* 40
4.1 Author rank plotted against the number of novels by that author 56
4.2 Bubble plot of writers by country, matching nationality to output 56
4.3 Publications per country by year 57
4.4 Cluster dendrogram of Harlequin Presents novels by year 59
4.5 Principal Components Analysis of Harlequin Presents, with years overlaid 60
4.6 Early period words in Harlequin Presents 61
4.7 Middle period words in Harlequin Presents 62
4.8 Late period ‘non-status’ words in Harlequin Presents 62
4.9 Late period ‘status’ words in Harlequin Presents 63
4.10 Number of authors publishing their first title and last title in the Harlequin Presents line 65
7.1 *Film Weekly* 1948–49 cover and page 1 100
7.2 *Film Weekly* data extracted to spreadsheet 101
7.3 Drive-ins and touring circuits in Victoria, 1955–56 103
7.4 Capri cinema, Goodwood 105
8.1 Thematic clusters in the Australian Twittersphere 120
8.2 Subdivisions within the politics cluster of the overall network map 121
8.3 Participation by Australian users in the hashtags #auspol (left) and #ausvotes (2010, right) 123
8.4 Participation by Australian users in the hashtags #masterchef (left) and #royalwedding (right) 124
8.5 Sharing of links to ABC News (left) and news.com.au (left) by Australian users 125
10.1 Network overview of the Flickr tag search ‘screenshot’ 142
10.2 Network overview of the tag ‘secondlife’ on Flickr 148

10.3 Network visualization of the tag 'worldofwarcraft' 150
10.4 Network visualization of the Flickr tag 'FPS' 153
10.5 Author's ego network of Flickr contacts 155
10.6 Network of connections of the most prolific poster in the 'game screenshots' Flickr user group 155
10.7 Visualization of the Flickr user Jokay's ego network of contacts and images commenters 156
10.8 Twitter network overview of user 'Jokay' 157
10.9 Zoomed view of the lower-center cluster of Figure 10.8 157
12.1 Australian novelists—number of novels written and number of 'works about' 179
12.2 Katharine Susannah Prichard: number of novels by and 'works about', 1900–2006 181
12.3 Total number of 'works about' 182
12.4 Number of unique authors in the yearly top ten for 'works about', per decade 183
12.5 Proportion of critical attention to top ten authors, per decade 185
12.6 Proportion of authors who feature in the top 50 for both academic and newspaper criticism, 1950–2006 187
12.7 Proportion of women authors in the top ten, 20, and 50 for academic attention, 1950s–2000s 188
12.8 Proportion of women authors in the top ten, 20, and 50 for newspaper attention, 1950s–2000s 188
15.1 Port Botany wharf data visualization 232
15.2 Port Botany road data visualization 234

Tables

3.1 Forty most significant keywords of the Dickens corpus as compared with the Wilkie Collins corpus (sorted according to the log-likelihood ratios [LLR]) 29
3.2 A result of running Random Forests 31
3.3 Important variables: Dickens markers and Collins markers (in order of importance) 32
3.4 Important V: Dickens versus the reference corpus (in order of importance) 33
3.5 Frequent 5-grams including 'eyes' in the Dickens corpus 35
3.6 Frequent 5-grams including 'eyes' in the eighteenth-century corpus 35
3.7 Frequent 5-grams including 'eyes' in the nineteenth-century corpus 36
3.8 Frequent 5-grams including 'hands' in the Dickens corpus 38
3.9 Frequent 5-grams including 'hands' in the eighteenth-century corpus 38

3.10 Frequent 5-grams including ‘hands’ in the nineteenth-century corpus 39
12.1 Properties of the dataset extracted from AustLit 178
12.2 Proportion of critical attention to top ten authors, per decade 186
13.1 Example of stepwise formalization from text to vector data 196

Contributors

Paul Longley Arthur is Professor of Digital Humanities, University of Western Sydney. He was previously Deputy Director, Centre for European Studies, a joint initiative of the European Commission and Australian National University. From 2010–13 he was Deputy Director, National Centre of Biography, Australian National University, and Deputy General Editor, *Australian Dictionary of Biography*.

Katherine Bode is Senior Lecturer in Literary and Textual Studies at the Centre for Digital Humanities Research at the Australian National University. Her most recent monograph, *Reading by Numbers: Recalibrating the Literary Field* (2012), explores the critical potential of quantitative book historical and bibliographical methods for literary history.

Axel Bruns is an Australian Research Council Future Fellow and Professor at Queensland University of Technology. He is the author of *Blogs, Wikipedia, Second Life and Beyond: From Production to Produsage* (2008) and *Gatewatching: Collaborative Online News Production* (2005). His current work develops new methodologies for studying public communication in social media.

Jean Burgess is Associate Professor of Digital Media Studies, Director of Research Training for the Creative Industries Faculty, and Deputy Director of the Australian Research Council Centre of Excellence for Creative Industries & Innovation at Queensland University of Technology. Her research focuses on social and mobile media platforms.

Mark Byron teaches and publishes across the genres and practices of Modernism as well as textual and editorial theory. His current work developing digital scholarly editions of complex Modernist texts and their manuscripts includes the *Watt* module of the Samuel Beckett Digital Manuscript Project. He is the author of *Ezra Pound's Eriugena* (2014).

Mark Coté is Programme Director of the Masters in Digital Culture and Society at King's College London. Mark writes on social media, digital culture, and the relationship between the human and technology. He is currently examining 'big social data' through an Arts and Humanities Research Council-funded research project with Tobias Blanke.

Øyvind Eide is a lecturer and research associate at the Chair of Digital Humanities, University of Passau, Germany. His research interests focus on the modelling of cultural heritage information, especially as a tool for critical engagement with the relationships between texts and maps as media of communication.

Jack Elliott is a doctoral candidate at the University of Newcastle. His dissertation is on machine learning and romance novels.

Julia Flanders is Professor of Practice in English and Director of the Digital Scholarship Group at Northeastern University. She directs the Women Writers Project and serves as editor in chief of *Digital Humanities Quarterly*. Her research focuses on digital text representation and editing, and on theories of work in digital humanities.

Tim Highfield is a Research Fellow at Queensland University of Technology and Curtin University. His research examines political and cultural uses of social media, including information flows around breaking news and popular culture. His research website is located at http://timhighfield.net/.

Alan Liu is Professor of English at the University of California, Santa Barbara. His books include *The Laws of Cool: Knowledge Work and the Culture of Information* (2004) and *Local Transcendence: Essays on Postmodern Historicism and the Database* (2008). He is founder of the 4Humanities.org advocacy initiative.

Willard McCarty is Professor in the Department of Digital Humanities, King's College London, and in the Digital Humanities Research Group, University of Western Sydney; Editor of *Interdisciplinary Science Reviews* (2008–); founding Editor of *Humanist* (1987–); and recipient of the Roberto Busa Award (2013), Alliance of Digital Humanities Organizations. See www.mccarty.org.uk/.

Richard Maltby is Professor of Screen Studies and Executive Dean of the Faculty of Education, Humanities and Law at Flinders University, South Australia. His publications include *Hollywood Cinema* (2nd ed. 2003),*'Film Europe' and 'Film America': Cinema, Commerce and Cultural Exchange, 1925–1939* (1999), and *Explorations in New Cinema History: Approaches and Case Studies* (2011).

Christopher Moore is Lecturer in Digital Media and Communication in the School of Arts, English and Media at the University of Wollongong. He is a researcher in games studies and is currently investigating the role of online personae within indie and independent cultures of games production.

Tara Murphy is Senior Lecturer in the School of Physics at the University of Sydney. She is currently leading the VAST project to investigate astronomical objects that vary on rapid timescales. Tara is also interested in interdisciplinary applications of novel computer science techniques to data-intensive research in astronomy and beyond.

Peter Robinson is Bateman Professor of English at the University of Saskatchewan. He is developer of the texual editing program Collate, used by many textual editing projects worldwide, and of the Anastasia electronic publishing system.

He is active in the development of standards for digital resources and tools for collaborative editing.

Ned Rossiter is Professor of Communication at the Institute for Culture and Society at the University of Western Sydney and teaches in the School of Humanities and Communication Arts. He is currently working on a collective project entitled *Logistical Worlds: Infrastructure, Software, Labour* (logisticalworlds.org).

Sydney J. Shep is Reader in Book History at Victoria University of Wellington and The Printer at VUW's Wai–te–ata Press : : Te Whare Tā o Wai–te–ata. She directs the Print Culture eResearch Hub which hosts 'The Printers' Web', the New Zealand Reading Experience Database, and the Digital Colenso, a prosopographical collaboratorium.

Tomoji Tabata is Associate Professor of Language Informatics at the Graduate School of Language and Culture, University of Osaka. His primary research interests are in forensic stylometry and corpus–stylistic investigation of the language of eighteenth- and nineteenth-century British fiction with special reference to Charles Dickens.

Paul Turnbull is Professor of Digital Humanities at the University of Tasmania, and an Honorary Professor of eHistory at the University of Queensland. His digital research projects include 'South Seas', a major online resource focused on James Cook's Pacific voyaging; The Gugu Badhun Oral History; the Digital Guide to Sources in North Queensland History; and PaperMiner: Spatial Mapping of Australian Historical Newspapers.

Deb Verhoeven is Chair and Professor of Media and Communication at Deakin University. She is a founding member of the Australasian Association for the Digital Humanities (aaDH), Project Director of the Humanities Networked Infrastructure initiative (HuNI), and a member of the Tasmanian government's Digital Futures Advisory Committee.

Dylan Walker is a doctoral student in screen studies at Flinders University, South Australia, researching film distribution and exhibition in rural South Australia during the 1930s. He has published articles on rural cinema exhibition and his book, *Adelaide's Silent Nights* (1995), deals with the history of cinemas in Adelaide, South Australia, during the silent era.

Mike Walsh is Associate Professor in the Screen and Media Department of Flinders University in Adelaide, Australia. He has published regularly on Australian cinema and is currently completing a history of the South Australian Film Corporation. He is also a programmer and writer for the Adelaide Film Festival.

1
Collecting Ourselves

Katherine Bode and Paul Longley Arthur

Digital humanities has become an influential and widely adopted term only in the past decade. Beyond the rapid multiplication of associations, centres, conferences, journals, projects, blogs, and tweets frequently used to signal this emergence, if anything characterizes the field during this time it is a concern with definition. This focus is acknowledged and reflected, for instance, in Matthew Gold's 2012 edited collection, *Debates in Digital Humanities*. The debates surveyed are overwhelmingly definitional: 'As digital humanities has received increasing attention and newfound cachet, its discourse has grown introspective and self-reflexive' (x). Questions that Gold identifies as central to and expressive of the emerging field include: Does one need to build or make things to be part of the digital humanities? 'Does DH need theory? Does it have a politics? Is it accessible to all members of the profession', or only those working at elite, well-funded institutions? 'Can it save the humanities? The university?' (xi).

The 2013 collection *Defining Digital Humanities: A Reader* (Terras et al. 2013) also reflects this focus, bringing together historical and contemporary readings on the act of defining digital humanities, many of which, not incidentally, are canonical in the field. Other areas of activity are equally self-reflexive, including the field's various manifestos[1] and the annual *Day of Digital Humanities*, where definitions are crowdsourced and participants are asked to document through text and image 'what digital humanists really do'. Despite this long-standing preoccupation, no clear agreement on a definition has emerged beyond broad references to research, teaching, and technical innovation at the intersection of humanities and computing. And within this broad description, commentators emphasize different aspects of the intersection—historical, institutional, political, economic, or social—as the aims and scope of digital humanities continue to be debated.

Why, then, is digital humanities so focused on defining itself, yet unable to arrive at an agreed-upon definition? Many have assessed this situation from a positive angle. Gold, for instance, suggests that such introspection simply marks 'a field in the midst of growing pains as its adherents expand from a small circle of like-minded scholars to a more heterogeneous set of practitioners who sometimes ask more disruptive questions' (Gold 2012, x–xi). Alan Liu (2013) identifies this focus as a characteristic that digital humanities shares with a number of past

fields, and thus, presumably, a relatively normal stage of development and maturation. We could add that even the most mature disciplines have adapted and shifted their boundaries in recent decades as the practices and rhetoric of interdisciplinary research have extended the scope of traditional pursuits. The definitional debate has also expressed many positive aims for digital humanities and the humanities more broadly, including openness beyond the university; the importance of interdisciplinary and global connections, conversations, and collaborations; critiques of established forms of hiring, peer review, and publication; and the importance of valuing—and making a case for the value of—humanities scholarship.

Yet the coin has a negative side, too. The focus on definition has fed into internecine and public battles about who is in and out of digital humanities—for example, whether one needs to code, or just to 'build', or neither, to be considered a 'digital humanist' (Ramsay 2011). Such disputes make the field appear cliquish and arguably occur at the expense of—or at the very least overshadow—the actual work of doing digital humanities and advancing the field by showing, rather than proposing or imagining, what can be achieved and discovered. This criticism could be pushed further, as a number of commentators within and outside digital humanities have done, to describe the work in digital humanities thus far as inadequate to constitute a field of humanities scholarship worth defining.

The most cogent and confronting expressions of this position come from within digital humanities itself. For instance, Patrick Juola's 2008 analysis of article citations and author affiliations in the flagship journal *Computers and the Humanities* (*CHum*) highlighted the 'minimal' impact of this research on mainstream humanities research. Building on this critique, in 2011 Andrew Prescott described digital humanities as too focused on internal debates—definitional, institutional, and technical—and displaying a lack of engagement with critical theory that left the field 'perilously out of touch with the modern study of the humanities' (69). The result, Prescott argues, is a 'collective failure to produce scholarship of outstanding importance and significance' (63). Liu's 2013 *PMLA* article encapsulated such criticisms as 'the meaning problem' in digital humanities: the seeming inability for research in this area to move from data (or textual models or visualizations) to arguments and interpretations that contribute to knowledge and debates in the broader humanities.

The fact that major figures in digital humanities question whether the field has made any real contribution is concerning enough; what makes these critiques even more worrying is how closely they mirror earlier criticisms and supposed solutions. Reflecting on the first 24 years of *CHum*, the same journal Juola analyses, in 1991 Rosanne Potter argued that most of its authors 'let the computer define what to look for, and the statistician define what was significant' while 'neglect[ing] reading what others were doing and had done'; this 'toddler and teenager thinking' produced a lot of counting, but not much in the way of insights relevant to mainstream humanities disciplines (427). A few years later, referring to this synopsis, Mark Olsen—like Prescott—diagnosed the urgent need to 'address the issues surrounding the general failure of our discipline to have a significant impact on the research community as a whole' (1993/1994, 309) by engaging with contemporary

theoretical insights and approaches.[2] The fact that the same problems—and purported solutions—remain despite more than two decades of work and a name change (from humanities computing) indicates the complexity of the 'meaning problem' for digital humanities.

A number of chapters in this collection continue and develop these and other critiques (see, for example, Bode and Murphy, Robinson, Rossiter, and Turnbull). But Willard McCarty's description of digital humanities' shortcomings is especially sobering, in part because his chapter was originally delivered as a lecture upon receiving the Busa Award, the most prestigious recognition of digital humanities scholarship offered by the Alliance of Digital Humanities Organizations:

> Make no mistake: we are *surrounded* by mature, subtle civilizations of enquiry, whose intellectual resources dwarf our own in volume, variety, and sophistication . . . . We need far more than the luck of the moment, dozens of sessions at the MLA, THATCamps everywhere, millions of tweets, thousands of blogs, and so on and so forth. We need *resonance* with the intellectual cultures of the arts and humanities . . . and [of] the techno-sciences.

McCarty points to fear—of our confrontation with 'the uncanny *otherness* of computing' (this volume, 294) and of our being unable to emerge unscathed from this encounter—to explain why digital humanities (and humanities computing before it) has not achieved the stature of other areas of inquiry.

These criticisms of digital humanities provide a potential answer to the question of why the field has been so focused on definition, and why agreed-upon definitions, despite this effort, have not emerged. Instead of occurring at the expense of, or overwhelming, the work of doing digital humanities, it may well be that definition has been pursued as a stand-in for the insights and discoveries the field is not providing. Pursuing this line of reasoning, perhaps the unprecedented attention that digital humanities has received in recent years accounts for the current fever pitch of definitional debate. Heralded as the 'next big thing' (Pannapacker 2009)—or even just 'the thing' (Pannapacker 2011)—by many within and outside the academy, and even put in the position of saving the humanities from its own crisis of meaning (Liu 2013), the focus on definition may be an attempt to provide something—anything—for those who look to the field for salvation at a time of institutional and financial pressure, and even epistemological and ontological crisis.

This collection was conceived in the context of such criticisms of digital humanities—with the aim of advancing the field beyond definitional debates to show, rather than describe, what digital humanities is, what it can do, the contribution it makes to humanities research, and the role it can play in the future: its research, theories, and methods. To attempt this ambitious contribution, we embarked on the collection with two main ideas. First, rather than predetermine the areas we believed to be part of digital humanities, we sought contributions from a wide range of scholars who had self-identified with the field by attending

the inaugural conference of the Australasian Association for Digital Humanities at the Australian National University in Canberra in 2012.

The result of this self-selecting approach is a collection that traverses some of the field's key fault lines. Indeed, quite unintentionally, the resulting collection demonstrates the broad version of digital humanities—its areas of need and promise—that Liu has been mapping out and advocating for a number of years (2011). Describing digital humanities from an ethnographic standpoint, Liu (2013) points to media studies—concerned with new media objects and networked, visual, and multimodal work—and traditional humanities computing—focused on technical questions and support—as neighbouring tribes to the core digital humanities, which arises mainly from traditional humanities disciplines and is predominantly concerned with textual materials as well as the examination and value of the old. This collection unites members of these 'neighbouring tribes' to demonstrate the importance of traditional and new scholarly objects and methods for digital humanities.

As is appropriate, given the emphasis internationally on literary studies in digital humanities research, that discipline provides one focus for this collection. However, such literary scholarship extends from established humanities computing areas of digital scholarly editing and stylistic analysis to book history and quantitative literary history. Moreover, these literary chapters sit alongside, and resonate productively with, a wide range of chapters relating to new media studies and theory, as well as contributions from researchers in film studies, history, cultural heritage, and even astrophysics. In addition, far from lacking the 'critical awareness of the larger social, economic, and cultural issues' at stake in humanities research (Liu 2011, 11)—a charge frequently levelled at digital humanities—this collection demonstrates a clear focus on relations of power and inequality, including their relation to gender, nationality, and global capital. Finally, while incorporating researchers from across the globe, the collection's origins give it an Australasian inflection, showcasing work from Australia, New Zealand, and Japan, and thus providing an important addition and corrective to the North American and European focus of many previous edited collections.

The second way we endeavoured to advance digital humanities beyond definition was to request contributions that did not merely describe digital humanities projects and methods but made an original contribution to research. The result is a collection of arguments, analyses, findings, and theories of relevance and value to multiple areas of the humanities. Such research has importance far beyond what it says about digital humanities, and we end this introduction by describing some of the specific ideas and arguments presented. Collected together, these chapters clearly demonstrate (rather than describe) the capacity of digital humanities to function as a site of rich conversations between and across disciplines.

What this collection also reveals, and what is particularly surprising given the diverse range of topics explored, is a set of significant commonalities in the approaches used and, more essentially, in the epistemological questions posed. Not all of these commonalities are new to our understanding of digital humanities, with themes such as the power of 'big data', the relationship of 'close' and 'distant'

reading, and thinking through building all much discussed in recent years. Yet deployed in ways that advance our understanding of the world and our place in it, these common themes begin to indicate some of the intellectual possibilities of digital humanities: that is, not only how digital humanities can contribute to humanities disciplines, but what it does differently from these disciplines, and why and how this difference is important in understanding and analysing human culture and society now and in the future.

The core commonality that emerges is the *centrality of collections* as a product of, resource for, object of, and epistemological challenge in digital humanities research. While the meaning of collections has expanded in the digital age—here including digital archives and libraries as well as collections of images, documents, words, and metadata—its centrality to the humanities is long-standing. As the authors of *Digital_Humanities* write,

> Collection-building and curation have remained constants of humanistic knowledge production from remote antiquity through early modern courts to the academics of the Baroque era to late nineteenth century universities where chairs were typically associated with research collections. These domains became disjointed from the mainstream of scholarly practice only during the late print era, and are once again becoming integral to many forms of Digital Humanities practice.
>
> (Lunenfeld et al. 2012, 32)

Rather than simply continuing a tradition, the chapters in this book exemplify and interrogate how digital collections are motivating new ways of doing research—and indeed, of thinking—in the humanities.

The *critical potential of data-driven analysis* to enable new perspectives and insights provides another common thread. In a wide range of topics—including Harlequin Romance fiction (Elliott), the operations of Australian cinemas (Maltby et al.), and the ‘digital human’ created in assemblage with the smartphone (Coté)—these chapters demonstrate the capacity of data-driven analysis to indicate patterns and conjunctions that could not otherwise be perceived, and which enable us to understand cultural phenomena in revealing and challenging ways. As well as seeing culture from a ‘distance’, a number of contributors integrate data-rich analysis with exploration of particular instances: whether this means using data to highlight phenomena that are explored further by ‘analogue’ means (such as interviews, archival research, or textual analysis), or moving iteratively between these two levels of analysis. In combining what has been called ‘close’ and ‘distant’ reading, such research brings together modes of analysis that are in some quarters seen as paradigmatic of a supposed opposition of humanities and digital humanities, and takes a step that Liu (2013) identifies as vital for solving the ‘meaning problem’ in digital humanities.

Another commonality across the collection is a focus on *thinking through building*. While also an established theme in digital humanities—developed in discussions of the role of knowledge representation and modelling in the field—chapters

in this collection realize the potential of this process by showcasing the intellectual, critical, and theoretical outcomes that thinking through building enables. The emphasis, in other words, is as much on thinking as on building, and these chapters highlight the iterative or dialogic movement between building and thinking, as models or prototypes prompt new questions and arguments that in turn motivate the creation of new models or prototypes, and so on. In his chapter in this collection, McCarty expresses ambivalence about the critical potential of modelling—a form of thinking through building that he himself did much to clarify—for digital humanities research, because it is 'unable to do more than work through consequences of interpretation that had already happened—elsewhere by other means' (this volume, 293). The work that follows shows the continuing importance and vitality of this approach, which in shifting our perceptions by degrees supports interpretations that would not be possible otherwise.

A key reason why the chapters in this collection succeed in using building to move beyond existing ways of thinking points to another commonality: *an explicitly self-conscious and critical approach to the nature and implications of collections*. As Julia Flanders explores in this volume, the current remediation of our cultural heritage into digital forms provides the opportunity to interrogate our assumptions about, and approaches to, collections before they become ingrained, normalized, and ultimately invisible. A number of chapters take up this challenge by highlighting the methodological and epistemological challenges and potential of analysing mediated objects, from mid-eighteenth-century documents used in the border negotiations between Denmark, Norway, Sweden and Finland (Eide), to the collective of Australian Twitter users (Bruns et al.) or national biographical datasets (Arthur). While Prescott criticizes digital humanities as atheoretical, even antitheory, if theory is 'a reasonably systematic reflection on our guiding assumptions' (Eagleton, cited in Prescott 2011, 68) then theorizing is precisely what the authors in this collection are doing. By interrogating how the organization of knowledge shapes what and how we can think—and using this awareness to imagine new modes of organization and, hence, new modes of thought—an explicitly self-conscious approach to collections manifests a dynamic mode of theoretical work.

A final commonality among chapters in this collection is a way of working that we term *ambitious generosity*. Ambition is clearly apparent in the scale and aims of these projects. Whether opening new areas for research or providing resources to reinvigorate and transform existing areas of scholarship, or the humanities as a whole, the authors in this book think big about the critical potential of digital techniques and technologies for humanities research. Much of this ambition relates to the focus on collections, which render these projects large in conception and membership, frequently involving not only a single researcher, or even a team of researchers, but multiple collaborators, many of whom will never meet or even know of each other, working on the same collection and thus contributing to the same broad agenda of expanding knowledge in a particular area or field.

What is remarkable about the ambition on display here is how closely and constitutively it is tied to generosity, a feature that is again intimately associated

with the centrality of collections to such research. While the collections that underpin many of these projects were created to pursue particular questions or interests, they are also intended to support and enable a much wider range of research areas and approaches—and to be used by others, including academics as well as citizen scholars beyond the academy. Indeed, many of the chapters in this volume deliberately blur the distinction between these two groups, whether by creating resources that refuse to privilege traditional academic forms of knowledge or by proposing systems that make such knowledge practices accessible and employable by all. In their ambition and their generosity, and in the dynamically theoretical modes of scholarship the following chapters demonstrate, the authors involved in the research presented here collectively position digital humanities as moving beyond a concern with itself to making a productive—and potentially transformative—contribution to our knowledge of the world and our place in it.

* * *

While united by these common threads and a focus on research findings, the collection is divided into four sections representative of particular fields of investigation or topics of analysis. The first of these—'Transforming Disciplines'—features digital research from the traditional humanities disciplines of literature and history. In 'Exercises in Battology', Mark Byron demonstrates how a digital edition of Samuel Beckett's manuscript *Watt*—itself part of a large-scale collaborative project to digitize the complete manuscripts of Beckett's published literary oeuvre—both extends the complexity of, and clarifies, this important work. As well as being necessitated by the extensiveness of the manuscript, a digital edition has important benefits for hermeneutic and textual criticism: for instance, Byron shows how moving between 'textual features at different orders of magnitude and in various modes' can provide insights—for instance, into characterization—not previously possible (this volume, 16).

Tomoji Tabata's 'Stylometry of Dickens's Language' also analyses a canonical literary figure—in this case, Charles Dickens. Tabata uses mining techniques for stylometric differentiation and forensic analysis of text to provide a foundation for deeper understanding of Dickens's language. Dickensian keywords and markers are contrasted with those in the works of his contemporary Wilkie Collins, and Dickens's works are compared with a large reference set of eighteenth- and nineteenth-century texts. Arguing against the value of culling techniques such as clustering, Tabata applies a machine learning approach using the Random Forests classification algorithm to show how Dickens's language can be distinguished and to identify the elements that constitute Dickensian style. Jack Elliott's analysis, 'Patterns and Trends in Harlequin Category Romance', moves us to the realm of mass-market fiction, using bibliometric and stylometric methods to provide a number of important insights into this industry, including its authorship and production practices. In particular, Elliott's analysis of changes in the titles of these novels challenges the long-standing and prevailing view of the industry

as static—offering a particular type of product and message to a specific type of reader—by identifying seismic shifts that significantly alter the narrative on offer.

In 'The Printers' Web', Sydney J. Shep describes the collaborative and interdisciplinary creation of a multidimensional digital biography of nineteenth-century printer Robert Coupland Harding, which is simultaneously an investigation into global print culture via the common practice of cut-and-paste journalism. Such journalism—presented as a predigital manifestation of open access as well as of viral texts and transnational readerships—is frequently described metaphorically as a network, web, or kaleidoscope. This project shows how digital methods and tools can be employed to investigate concrete manifestations of these hitherto metaphoric concepts by exploring the movement of texts through space and time, as well as the operations of a specifically printing-based culture and the life of an important historical figure. Paul Longley Arthur's chapter, 'Biographical Dictionaries in the Digital Era', approaches biography from a different perspective, reflecting on the complex transition from print to digital of a major national biographical dictionary and its reconceptualization as a virtual environment for social history and network analysis. This case study refers to historical, cultural, and technical factors to demonstrate how digital approaches have enabled a shift of emphasis away from the individual, focusing instead on tracing networks of associated lives—and in the process opening up new questions for biographical scholarship and extending the conventional role of biographical reference works in the digital environment.

The second section, 'Media Methods', presents digital humanities investigations into various forms and practices of media, both new and established. In 'Digital Methods in New Cinema History', Richard Maltby, Dylan Walker, and Mike Walsh describe and demonstrate a transition from film studies to new cinema history, where analysis of quantitative data shifts the focus from individual films, conceived as texts, to cinematic events and their circulation and consumption in local and global markets. Focusing on the AusCinemas database, the authors challenge prevailing assumptions about the history of cinema—for instance, the cause of cinema closure in Australia and that market's relationship with America—while also describing their experiments with crowdsourcing—their successes and failures—as a means of progressing such research and moving it beyond the academy.

The critical potential of data-rich analysis is also key to 'A "Big Data" Approach to Mapping the Australian Twittersphere' by Axel Bruns, Jean Burgess, and Tim Highfield, which takes a new approach to analysing Twitter, moving beyond discrete cases and investigation of particular topics and events via hashtag studies to explore the larger networked context in which this communication occurs. As the first attempt to identify, extract, analyse, and visualize the breadth and depth of a national portion of the Twittersphere, this study both explores key methodological challenges faced by large-scale analysis of proprietary systems and—in showing the distribution of particular topics across a wider network—analyses the operations of the Twittersphere and formation of social groups and identities online.

In 'iResearch', Mark Coté reports on an experiment that captures big social data (BSD) on mobile social media use in everyday interaction and communication. He argues that the data-constituted digital human—defined as 'a methodological covalence in the BSD database' and in terms of 'the human as always already in a constitutive relationship with technology' (this volume, 137)—should be a vital object of study in digital humanities. In particular, as we become 'native speakers' of technological syntax, BSD emerges as a cipher or a mediation of the digital human, thus providing a key area for understanding contemporary cultural heritage, society, and 'the structural complexity of the digital human and its cultural forms' (this volume, 132).

In 'Screenshots as Virtual Photography', Christopher Moore uses visualizations of the movement of video game screenshots—'a ubiquitous form of visual communication' (this publication, 141)—across social media platforms to investigate online identity formation and curation. The questions prompted by this process motivate new visualizations, investigations, and questions in a dialogic process that uses visual modelling as a basis for theorizing. Moore arrives at a conception of screenshots that challenges traditional hermeneutic and phenomenological interpretations derived from photography and instead highlights the performativity, materiality, affect, and motility of this form of online communication.

While apparent throughout the book, an explicitly self-conscious, critical approach to collections—whether of metadata, images, documents, or words—is foregrounded in the third section, 'Critical Curation'. In 'Rethinking Collections', Julia Flanders uses the language of modelling to situate the creation and coherence of collections not as neutral or standardized practices but as activities laden with specific intentions and driven by specific questions. Identifying the network and the patchwork as two dominant models for collections, Flanders considers the epistemological implications for the digital humanities of their different *modus operandi*: the former emphasizing connection, seamlessness, total commensurability, and increasing harmonization; the latter, assemblage, negotiation, and transaction.

This networked view of collections has the potential to support the positivist view of data—as truth or fact—that Katherine Bode and Tara Murphy critique in their chapter, 'Methods and Canons'. These authors analyse critical attention paid to Australian novelists and the issues of canon formation and cultural value relating to such judgements. However, to avoid the common association of data analysis with truth and comprehensiveness, Bode and Murphy publish their data along with a detailed account of the assumptions, arguments, and values that were involved in its creation. The approach they model emphasizes data's status as outcome, rather than unquestioned and unquestionable basis, of argument.

In 'Reading the Text, Walking the Terrain, Following the Map', Øyvind Eide likewise highlights the distance between, rather than equivalence of, data and the material world. Specifically, he employs computer modelling of maps and texts to identify and investigate the limitations of each as modes of representing the

landscape. As Eide notes, all humanities scholars are aware that every medium has the capacity to represent only certain aspects of reality. However, modelling provides an enhanced understanding of both the nature of representation and mediation and how such systems play out at the micro level.

In 'Doing the Sheep Good', Deb Verhoeven investigates the role of databases in contemporary scholarly practices and knowledge formation. Far from being a neutral intervention, the 'production, organization, and communication of data' (this volume, 207) can encode its own implicit value system, fundamentally influencing how data is used and reused and the resultant conclusions that can be drawn. The author refers to case studies that link digital humanities with creative arts research, within and outside the academy, arguing for deeper and extended forms of engagement between modern academic work and wider publics.

In 'Materialities of Software', Ned Rossiter analyses the ways in which computational parameters increasingly mediate the production of knowledge, both in digital humanities itself and in the interests of global capital. According to Rossiter, existing computational methods—graphs, charts, digital maps, and so on—operate within and therefore have difficulty perceiving, let alone critiquing, the self-referentiality of data. In making a case for the need to invent new methods—relating to graphic design, game development, and digital visualization—Rossiter posits a mode of data analysis that resonates with Flanders's suggestion of a patchworked approach to collections: specifically, one that focuses not on equivalences or similarities—which 'integrate and make uniform data' in ways that erase geocultural differences—but on 'sites . . . where interoperability breaks down' as data rubs up against political, social, geographical, and material disruptions (this volume, 228).

In the book's final section, 'Research Futures', scholars who have for many years worked at the intersection of humanities and computing reflect on failings of the past to propose directions for the future. Peter Robinson's 'Digital Humanities' criticizes the dominant current model for digital humanities research—focused in centres and dependent on external funding—as unsustainable and argues instead for a future where an underlying infrastructure enables individual academics as well as scholar citizens to create and enrich materials over the Web. In 'Margins, Mainstreams and the Mission of Digital Humanities', Paul Turnbull asserts the need to challenge ingrained cultural mistrust of digital resources and technologies, while questioning the extent to which digitally based humanities research should be seen as distinct from the wider activities of humanities research that are increasingly informed by, and often integrally reliant on, the use of digital media and technology.

The major challenges and opportunities for digital humanities in the future are also the focus of the final two chapters, both of which have been published previously, Alan Liu's in French online journal *Ina*, and Willard McCarty's in *LLC: The Journal of Digital Scholarship in the Humanities*. In 'The Big Bang of Online Reading', Liu identifies online reading as equivalent for the digital humanities, and the humanities more broadly, to the 'big bang' in physics: an event unleashing 'a multiplicity of forces, materialities, forms, and dimensions' (this volume, 275). Liu

argues that the resulting reconfigurations—of the media; of materiality; of literacy, in its sensory and social manifestations; and of cognition—together constitute a new research object in relation to which the digital humanities can offer important approaches and interventions.

In 'Getting There from Here', McCarty argues that the major challenge for digital humanities in the future is to realize we cannot avoid, and must instead embrace, being changed by the otherness of computing. Resting in—rather than retreating from—'the posthumanizing juncture where computing meets the humanities' will not divorce us from ourselves but help us to discover 'what it means to be human' (this volume, 306). By showcasing new approaches that harness not only the power of data but of our own humanity—in various forms of collectivity and ambitious generosity—this volume signals some of the ways in which digital humanities can work at this intersection, advancing discussion beyond a concern with definition to provide methodologically critical and dynamically theoretical modes of humanities research.

Notes

1. See, for example, *A Digital Humanities Manifesto* (Manifesto) and its more recent manifestation, *A Digital Humanities Manifesto 2.0* (Manifesto 2.0), as well as *Digital_Humanities* (Lunenfeld et al.).
2. Similar arguments and observations had been outlined by Olsen (1991) and Kenny (1992).

References

Gold, M. K. (2012). 'Introduction: The Digital Humanities Moment'. In *Debates in the Digital Humanities*, ed. M. K. Gold, ix–xvi. Minneapolis: University of Minnesota Press.

Juola, P. (2008). 'Killer Applications in Digital Humanities'. *Literary and Linguistic Computing* 23, no. 1: 73–83.

Kenny, Anthony. (1992). *Computers and the Humanities*. Ninth British Library Research Lecture. British Library, London.

Liu, A. (2011). 'The State of the Digital Humanities: A Report and a Critique'. *Arts and Humanities in Higher Education* 11, nos. 1–2: 8–41.

Liu, A. (2013). 'The Meaning of the Digital Humanities'. *PMLA* 128, no. 2: 409–23.

Lunenfeld, P., A. Burdick, J. Drucker, T. Presner, and J. Schnapp. (2012). *Digital_Humanities*. Cambridge, MA: MIT Press.

Manifesto. (2008). *A Digital Humanities Manifesto*. Available at: http://manifesto.humanities.ucla.edu.

Manifesto 2.0. (2009). *A Digital Humanities Manifesto 2.0*. Available at: http://manifesto.humanities.ucla.edu.

Olsen, Mark. (1991). 'What Can and Cannot Be Done With Electronic Text in Historical and Literary Research'. Paper for 'Modeling Literary Research Methods by Computer' session, Modern Language Association Annual Meeting, San Francisco, December.

Olsen, Mark. (1993/1994). 'Signs, Symbols and Discourses: A New Direction for Computer-Aided Literary Studies'. *Computers and the Humanities* 27, nos. 5–6: 309–14.

Pannapacker, W. (2009). 'The MLA and the Digital Humanities'. *Chronicle of Higher Education*, blog post, 28 December. Available at: http://chronicle.com/blogPost/The-MLAthe-Digital/19468/.

Pannapacker, W. (2011). 'Digital Humanities Triumphant'. *Chronicle of Higher Education*, blog post, 8 January. Available at: http://chronicle.com/blogs/brainstorm/pannapacker-at-mla-digital-humanities-triumphant/30915.

Potter, R. G. (1991). 'Statistical Analysis of Literature: A Retrospective on Computers and the Humanities, 1966–1990'. *Computers and the Humanities* 25, no. 6: 401–29.

Prescott, A. (2011). 'Consumers, Creators or Commentators? Problems of Audience and Mission in Digital Humanities'. *Arts and Humanities in Higher Education* 11, nos. 1–2: 61–75.

Ramsay, S. (2011). 'On Building'. Available at: http://lenz.unl.edu.

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

Part I

Transforming Disciplines

2

Exercises in Battology

Digitizing Samuel Beckett's *Watt*

Mark Byron

Literary scholarship on the work of Samuel Beckett is cresting a wave brought about by the new availability of significant primary material—particularly the notebooks in which Beckett recorded his reading notes and early fragments of literary composition (Nixon and Van Hulle 2013; Nixon 2011; Feldman 2006b; Pilling 1992, 1999; Maxwell 2006; Bryden et al. 1998), as well as published volumes of letters (Beckett 2009, 2012). This heightened documentary awareness in Beckett studies has stimulated renewed attention to such hermeneutic matters as text structure, continuities of themes and tropes in Beckett's reading and note-taking, and the varieties of citation and allusion in his texts. Analogous forms of literary scholarship and critical interpretation also flourish, including a series of monumental annotative studies of specific Beckett texts (Pilling 2004; Ackerley 1998, 2005; Ackerley and Gontarski 2004). Consequently this activity has provoked new insights into aspects of Beckett's composition processes and the material state of his manuscripts and published works. Manuscript documents in particular provide a broader and richer framework within which to describe the Beckett 'text' as a literary event or process. Current efforts to digitize these documents and provide authoritative transcriptions of them compel acute reflections on the status of Beckett's texts and the editorial methods adequate to the task of establishing and representing them in scholarly editions. The narrator of Samuel Beckett's 1953 novel, *Watt*, neatly captures this tension between diachrony and formal repetition in narrative form and textual structure when he states: 'Watt's sense of chronology was strong, in a way, and his dislike for battology was very strong' (Beckett 1959, 165). Such an observation resonates alongside attempts to conceptualize and mark up complex narratives for digital presentation in scholarly editions. Textual criticism and hermeneutics have still to catch up with some of the formal and structural experiments of Beckett's texts.

Of the richly various experimental literary works of Modernism, *Watt* enjoys a singular reputation for its complex, fragmented narrative structure. Its emblematic status as an avant-garde challenge to literary convention is manifested in its many footnotes, musical scores, 'manuscript' perforations in the narrative, and extended prose sequences of 'battological' prose. This textual-material endowment

is clarified further when the novel's very extensive manuscript archive is taken into account. This repository comprises nearly a thousand pages of drafts, revisions, and substantial narrative divergences from the published text. The manuscripts are famous in their own right as aesthetic objects—they are densely illustrated with Beckett's 'doodlings'—but very few scholars have read or studied them at any length, preliminary to any attempt to measure their evolution against published editions of the novel. The logic of a digital edition of this complex manuscript material is clear: to provide coherent ways of mapping the network of confluences and divergences between manuscript and published text, as a first, critical step towards exploring their semantic and hermeneutic implications. The sheer size of the manuscript record provides sufficient reason for digital representation, but a well-designed digital framework can go further: it can illuminate patterns and complexities in the material not otherwise readily evident in analogue form, presenting textual features at different orders of magnitude and in various modes (visual, statistical, diagrammatic, linguistic, and so on). This foundation opens up possibilities for further hermeneutic analysis and textual criticism, extending and challenging the ways in which this text and its manuscript have been understood both conceptually and as material objects.

The digital edition of the *Watt* manuscripts constitutes an early module in the Samuel Beckett Digital Manuscript Project (BDMP)—a large, international project that aims to digitize the complete manuscripts of Beckett's published literary works (Van Hulle 2011). Current work on the *Watt* module already suggests new modes of viewing and understanding the novel's narrative. The complex and often very subtle relationships between elements of the manuscript text—many of which do not appear in full or at all in the published text—are eminently conducive to digital display. The digital edition provides the careful reader with a more precisely calibrated appreciation for the reticulated networks of reference, association, and narrative lines in the manuscript material, and the implications of their recycled, revised, or omitted status in the published text. Such a digital remediation of archival documents provides benefits beyond the practical reach of print-based editions: the ability to visualize complex networks of citation and association across a large manuscript brings schematic focus to an otherwise bewildering document. Features such as multiple-page views aid in visualizing phases of composition, where doodles on verso pages, for example, often correlate to fluent, fractured, or otherwise distended prose on facing recto pages. XML-encoded information regarding patterns of emendation (the location of inserted material on the page, the nature of deletions, patterns of differently coloured inks, and so on) provides the tools by which to quantify, collate and interpret highly imbricated narrative material.

What kinds of benefits might accrue to the reader of such an edition? *Watt* is Beckett's most perplexing novel, presenting a turbulent text surface and a series of reticulated narrative digressions and redundancies. Consequently it has not received the same intensive critical attention as his other novels, despite its pivotal function as the transitional phase between the author's youthful, Joycean

displays of learning (*More Pricks Than Kicks, Murphy*) and the mature investigations into subjectivity and identity (*Molloy, Malone Dies, The Unnamable*) for which he became famous. *Watt* also marks the shift in Beckett's language of composition, from English to French. Although written in English, the manuscript contains numerous Gallicisms and other grammatical and syntactic tendencies bearing a French linguistic imprimatur. The obvious merits of the *Watt* digital manuscript consist in providing mediated access to complex primary materials, opening up new zones of hermeneutic inquiry and textual scholarship in Beckett studies. But the potential benefits and consequences of the digital edition extend further: to the context of Modernist scholarship more generally (particularly the famous 'documentary turn' of recent years), to the theory and practice of textual scholarship, and to the ways in which these fields each negotiate the potential challenges and opportunities of the digital sphere.

By digitizing the eccentric manuscript of an elliptical, avant-garde text, the structure and content of its physical archive may be better understood, to a degree equalled only by a direct appraisal of the physical documents, offering a schematic account of their codicological and semantic features. A clearer understanding of the fractured and fragmented nature of the composition also situates *Watt* more precisely with regard to Modernist experimentalism and literary fragmentation. The literary fragment has, in a sense, always been a central fact of textual scholarship: from the preservation of ancient literary and biblical texts in papyrus fragments (such as the Nag Hammadi Library, or Oxford University's online Oxyrhynchus Papyrus project at http://www.papyrology.ox.ac.uk/POxy/), to the complementary discourses of fragment and ruin that shaped the contest between Classicism and Romanticism in the eighteenth century (McFarland 1981; Seyhan 1992; Harries 1994), to the disintegrations and dismantlings of literary forms and genres in the twentieth-century transatlantic avant-garde. The *Watt* digital edition provides a model for how these fragments, documents, and texts relate to one another, but also raises questions of how they calibrate with larger modes of experimentalism in Modernist aesthetics, and how textual scholarship might respond to these basic challenges to the literary artefact in adequate scholarly editions.

The BDMP is an important experiment in evaluating the relative value and costs of digital tools and methods in scholarly literary projects more generally. It is an evolving case study in how such digital innovations might function with traditional scholarly methods (such as stemmatics and philology), as well as ways in which newer methods are being developed (such as genetic editing techniques and corresponding Text Encoding Initiative [TEI] protocols). The enormous potential for digital literary scholarship dwells in large part in the scale-changing implications of large-batch processing of information—although this notion of rendering the literary artefact into raw data can catalyse scepticism towards large-scale (and costly) computer-aided humanities research. The present task is to strengthen a middle ground in scholarly practice and digital design, where digital humanities and traditional scholarly methods might interact productively and thus extend their respective zones of capability. In his review of *The Evolution of*

Texts (Macé et al. 2006), Tuomas Heikillä takes stock of the rapid advances in digital methods in textual scholarship:

> There has hardly ever been a period during which textual scholarship (or philology) has seen so many new challenges and opportunities as in the past few years. The computerization of the study of texts—be they ancient, medieval, or modern—the use of digital images, and the transformation of the concept of 'edition' have all played their part.
>
> (2007, 298)

Heikillä observes how computer-aided approaches to stemmatology reconfigure the theoretical and empirical landscape. He delineates the virtues of cross-disciplinary experimentation in developing new stemmatic and philological tools, such as phylogenetic methods adapted from biology and information theory. But these intersections work most effectively when digital methods are directed by critical insight: Heikillä notes that in one experiment described in *The Evolution of Texts*—in which a range of methodologies were deployed to attempt the generation of a stemmatic record of an artificial manuscript tradition—the obvious utility of digital tools when applied to very large datasets did not eclipse the accuracy and elegance of classical methods in stemmatology. A critically reflective digital (or analogue) edition is predicated on elegance of design, utility of function, and the explanatory power of its conceptual models, whether cutting-edge, classical, or hybrid.

Reading with the *Watt* digital edition

The *Watt* digital manuscript edition aims to provide a mediated primary resource: the fullest representation of the *Watt* manuscript notebooks, housed at the Harry Ransom Center at the University of Texas at Austin, made accessible by way of an intuitive web interface (beckettarchive.org). The digital reproduction of the manuscript may be visualized in several ways—thumbnail images, parallel image-text functions, close-up images—and interacts with the XML-encoded transcription via a scrollover transcription function. The many doodles are presented schematically, recording their relationships with corresponding text material. These features allow for a more nuanced reading experience of the manuscript and illustrate how digital mediation can facilitate complex aspects of philology: stemmatic relationships between elements of a single document; the architecture of a literary object within a preliminary and evolving material form; and the complex interactions between word and image, often spatially disconnected or obliquely linked conceptually, otherwise invisible to all but the most assiduous codicologist. Examples of the word-image relation are abundant. Many of Beckett's doodles comprise humanoid figures, directly reflecting the evolution of the concept of character in the manuscript, and geometric figures visualize the permutative logic of several extended narrative phases.

The nearly 1,000 pages of Beckett's manuscript were composed from 1941 to 1945, mostly in the town of Roussillon in Vichy France following Beckett's abrupt