

Advancing Digital Humanities

Research, Methods, Theories

Bode Katherine; Arthur Paul Longley

ISBN: 9781137337016

DOI: 10.1057/9781137337016

Palgrave Macmillan

Please respect intellectual property rights

This material is copyright and its use is restricted by our standard site license terms and conditions (see http://www.palgraveconnect.com/pc/connect/info/terms_conditions.html). If you plan to copy, distribute or share in any format including, for the avoidance of doubt, posting on websites, you need the express prior permission of Palgrave Macmillan. To request permission please contact rights@palgrave.com.

14

Doing the Sheep Good

Facilitating Engagement in Digital Humanities and Creative Arts Research

Deb Verhoeven

But metaphor is never innocent. It orients research and fixes results.

Derrida 1980, 17

In the mid-1960s, with support from the National Science Foundation, Sol Worth (a communications scholar and documentary filmmaker) and John Adair (an anthropologist) took the unprecedented step of providing movie cameras to the Navajo community they were studying. This act is now regarded as one of the pivotal moments in the development of visual anthropology and more specifically the emergence of 'participant visual media research'. Worth and Adair hoped to glean new insights into Navajo culture through formal and thematic analyses of the films produced by members of the Pine Springs Reservation community. From the specific representations made, they expected to deduce defining cultural differences between themselves and the community under scrutiny.

Senior members of the community, in particular Sam Yazzie, were not entirely convinced by the proposed exercise and questioned Worth and Adair in some detail:

After some thought, Sam turned to Worth and through the interpreter asked, 'Will making movies do the sheep any harm?' Worth was happy to explain that as far as he knew no harm would befall the sheep if movies were made in the community.

Sam thought for a few seconds, and looking straight at Worth asked, 'Will it do them any good?' Worth was forced to reply that as far as he knew it wouldn't do the sheep any good.

Sam looked at us both and said ...

'Then why make movies?'

(Worth and Adair 1972, 4)

Reflecting on this exchange, Worth and Adair credit Yazzie's enquiry with an unwitting metaphorical dimension, presuming that his real anxiety was not the sheep but the Navajo community and 'how the new method of communication

that we were to teach his people could help the Navajo. How would making films support their values and their way of life?' (Worth and Adair 1972, 6). In contrast to their prosaic depiction of Yazzie's concern for community uplift, they propose their own theoretical and analytic ambitions: teaching a small group the skills of filmmaking in order to 'get away from an examination of man as an object and try to learn more about him as a subject' (Worth and Adair 1972, 26). For Worth and Adair the community cameras were a methodological ploy designed to create the ideal circumstances through which the ethnographers could critically assess the Navajo's creative facility with a new media technology.

Elsewhere I have written about the failure of the two anthropologists to concede agency for the nonhuman in this exchange (Verhoeven 2006). But what also remains unexplored in Worth and Adair's evaluation of this encounter is Sam Yazzie's deeper questioning of agency in research activities. By asking whether the filmmaking will harm or help the sheep, Yazzie suggests he is aware that the proposed ethnographic experiment is a metonymy for the research process itself. At the heart of his ovine enquiry is a question about whether the technology offered to the Navajo community (which pointedly includes sheep) might also provide opportunities for a co-production—whether the cameras might prove, for example, to be a diagnostic tool for the Navajo rather than simply an innovative exercise in content provision for the entertainment and advancement of the anthropologists.

Like Yazzie, this essay is also concerned with questions of inclusion and participation in the scholarly use of technologies, with the changing alignments of agency and the nonhuman made possible in contemporary research practices. In particular it explores the use of databases in academic research—the production, organization, and communication of data—and their consequences for the configuration and character of recent humanities and creative arts scholarship. Databases are a constitutive feature of modern scholarship and its administration, yet are still developing as objects of scholarly interest in their own right. The resulting emergence of political, practical, and philosophical questions about databases (around agency and automation, or privacy and transparency, for example) parallel growing public, commercial, academic, and government reliance on databases across a full range of social phenomena. Specifically, how and what do research databases 'do'? What positions might databases assume in the production of knowledge? How have changes to the nature of data itself also changed databases and their roles? How do debates about databases implicate our own practices as digital humanities and creative arts scholars? I explore these questions in detail through two case studies that exemplify some of the ways in which digital research technologies might bear on the practices of contemporary research scholarship (and vice versa).

Introduction

And such is the flood that even things that might have done good lose all their goodness.

Erasmus, *Adages* 2.1.1

Familiar accounts of information behaviour reach for easy metaphor. We live, it is said, in an age of information overflow (Gleick 2011). Information arrives torrentially and ceaselessly. Data courses without care through contemporary culture, simultaneously lifting us from the shallows of our former existence and drowning everything in the depths of insignificance. Shouldering aside the shaky structures of previous knowledge architectures, data overwhelms meaning in its wake. This pervasive story of data's relentless ascent is rendered in the style of the classics, as an apocalyptic 'deluge myth' of destruction and rebirth.

Databases, we might then infer, are the product of diluvian times. In an age of data inundation, databases are designed to harness, channel, and sometimes hinder the flow of information. They are conduits and strainers in the face of unreserved turbulence. Like weirs and levees, they attempt to govern the behaviour of information, enabling flow but limiting the damaging aspects of a current too strong. Databases render the flooded rivers of data navigable as well as facilitating the measurement of their waters. In their efforts, databases can also produce an impact on surrounding ecologies, creating downstream problems and accumulating collections of redundant or unwanted debris.

In quantitative terms alone, the metaphor of flooding, certainly as we have experienced it, may not be enough to even begin to describe the rising tide of the dataverse. In a recent study, the IDC predicted there will be some 40 zettabytes of data by 2020 (IDC 2012). By comparison, there are only 1.3 zettalitres of water stored in all the earth's oceans (National Oceanic and Atmospheric Association 2014). And the metaphor has further limitations. In this hydrologic economy of information, data is accorded an *a priori* status in which human effort is directed to acts of 'external forcing', to the (ultimately futile) imposition of boundary conditions (such as databases). Perhaps we might better aim to think about databases as operating in a type of socio-hydrology (Di Baldassarre et al. 2013), as a dynamic co-evolution of the interaction between humans and information systems. In this view, information is no longer autonomous. Instead, we recognize the ways we have opened up, the ways we think and live beyond data, as we also live with data. This alternative account of information inundation has implications that swell beyond the banks of singular research queries, washing away at the edges of disciplinary jurisdictions and the bulwarks of institutional authority in the academy, the state, and industry.

Saying that databases have proved to be especially useful for digital humanities and creative arts scholars across a wide range of research disciplines is almost glib. By organizing and structuring information, databases expand our capacity to comprehend and compare social and cultural phenomena in ways hitherto unimaginable. *But databases are not (just) a system for ideas. They are also an idea for a system.* As Kenneth M. Price points out:

A database is not an undifferentiated sea of information out of which structure emerges. Argument is always there from the beginning in how those constructing a database choose to categorize information—the initial understanding of the materials governs how more fine-grained views will appear

because of the way the objects of attention are shaped by divisions and subdivisions within the database. The process of database creation is not neutral, nor should it be.

(Price 2009)

The creation of databases requires considerable work; even off-the-shelf systems that encourage us to take the work of databases for granted rely on expertise, discipline, coordination, and large-scale resources to establish their infrastructure. Databases both result from and produce acts of communication and agency whose meaning and functions cannot be reduced to a narrow instrumentality. Databases are productions, and they are also productive and generative. For example, databases produce reactions; they intervene in the worlds they describe and propose. And databases both produce and certify knowledge. They also define (and redefine), through processes of simplification, classification, comparison, inference, and calculation, the relationship between those who evaluate and those who are evaluated. Technologies alone do not produce 'social good' any more than they produce social 'goods'. That is to say, databases don't just describe cultural information. Databases are cultural information, shaped by our organizational and scholarly needs and in turn shaping them. Constantly.

Constantly, because as the nature of data changes, so too does the nature of data collection and organization, and then so too do our strategies as scholars. The volume, velocity, and variety of contemporary data, for example, have changed the way databases operate. In the past it might have made sense to 'contain' data inside a database in order to 'clean' (normalize) and curate (give schematic structure to) information. Now, especially since the advent of 'the Internet of things', data arrives in clusters with predefined identity and versioning. Consequently there is a shift from targeted, discrete forms of information collection to always-on, ubiquitous, rapidly expanding, and accelerating data collection, which has resulted in significant changes to our understandings of information processing. New database formats such as NoSQL (not only SQL) and object-oriented databases, for example, have emerged to deal with this new type of data behaviour and rest on new theoretical organizing principles such as Identity and Versions, Attribution, Inferences, and Inheritances (in both the technical sense and also in the sense that any new theory of data must include precursor database taxonomies and technologies).

Some of these observations of the generative nature of databases have been taken up in emerging areas of study, such as critical code studies (Mackenzie 2006; Kitchin and Dodge 2011), software studies (Manovich 2001; Azuma 2009; Fuller 2008), platform studies (Montfort and Bogost 2009; Jones and Thiruvathukal 2012; Maher 2012), and other areas of the digital humanities (Folsom 2007; Hayles 2012; Liu 2008). This chapter complements these studies by specifically asking how established research behaviours such as authorizing, legislating, interpreting, and agenda setting are being challenged, unfolded, and reinvented by the technological and social changes brought about by new forms of digital media, including databases.

With this question in mind, Martin Weller's exploratory account of new forms of digital scholarship makes a case for four alternative 'scholarly functions' to the standard Boyer model of scholarship based on discovery, application, integration, and teaching. Weller proposes instead the key activities of engagement, experimentation, reflection, and sharing (Weller 2011). This essay presents two case studies in order to reflect on aspects of best practice for 'engagement' in digital humanities and creative arts research. Several key features are explored through these two case studies, including the ways digital humanities and creative arts scholars might:

1. *Span and expand (disciplinary) spaces* in ways that go further than the adoption of modular research practices, such as assembling scalable interdisciplinary academic teams to focus on problem-oriented projects, and instead invite us to consider research engagement in the broadest sense. For instance, to date very little critical debate has taken place about those cultural research methods that exist outside the academy and how powerful these might be in scholarly research. As Mike Savage suggests, we now have an opportunity, 'to broaden our repertoire and recognise the changing stakes involved in the circuits of "knowing capitalism"' (Savage 2009, 249).
2. *Emphasize making and maintaining relations*, both in terms of the data we work with and our sociality as academics (and not just with other academics). Carolyn Ellis reminds us that academic prosociality needs to go beyond interdisciplinarity. For Ellis, this means researchers who exercise a 'relational ethics' value and respect the connection between themselves and the people they study, and also between researchers and the communities in which they live and work (Ellis 2007, 4).
3. *Work at circumventing intermediaries* both by directly opening access to information to those outside the academy, and emphasizing high levels of transparency in academic practice. Thomas Osborne's powerful reconfiguration of the contemporary academic as a 'mediator' (Osborne 2004), responsible for progressing and moving knowledge between institutions, is tested by an emerging research ecology that gives both the public and scholars direct access to one another's ideas. But now we can equally imagine what the role of an academic 'disintermediator' might look like—producing opportunities for just-in-time knowledge via generous discovery interfaces, for example, or adopting the use of iterative project management tools that facilitate open processes of adjustment and improvement to multiple research stakeholders.
4. *Share and make available processes, materials, and knowledge* whilst being mindful of their meaning and value as a potential contribution for further research discovery, retrieval, exchange, reuse, and preservation. This includes understanding that our own digital traces, our successes and failures as we develop our research, are also retrievable and analysable.
5. *Commit to co-producing contents, technologies, infrastructures, and analyses* by recognizing how digital and social media platforms have altered traditional modes of production, including those of scholarly researchers themselves. In

this sense, 'engagement' means more than ensuring that the outcomes of our knowledge add value beyond our specific domains by improving access for others, including those outside the academy, to make use of it. Equally it also means creating research that is specifically designed to enable the co-production of knowledge with nonacademic groups through a commitment to multimodal knowledge. This could involve working in modes that are not necessarily written, for example (Bryson 2004). David Beer has considered how this gesture, coupled with changes in the behaviour of data itself, could prompt the profound rethinking of our understanding of the centrality of human agency in research: 'This is to technologise our research practices in a way that was not previously possible, in so doing it moves some more of the analytical processes of . . . research into the hands of machines' (Beer 2012).

The following two case studies are offered as partial explorations of these emerging principles of digital research practice. Whilst they describe vastly different registers of digital humanities and creative arts research, they share a mutual commitment to reflecting and following up on the changing relationship between scholarly researchers and their many stakeholders.

Case study 1

The Humanities Networked Infrastructure (HuNI) Project: Ontologies from Below

The Humanities Networked Infrastructure (HuNI) project (<http://huni.net.au>) is a major new infrastructure service for humanities researchers in Australia developed by a consortium of 13 institutions. HuNI ingests and aggregates data from a total of 31 different Australian datasets that cover a wide range of disciplines in the humanities and creative arts, including literature, biography, performing and visual arts, media studies, and linguistics. Through its use and development of innovative technologies and techniques, the HuNI project proposes some large questions, far beyond the specific queries of participating researchers: how, for example, might the opportunities presented by an unprecedented proliferation of networked data also challenge the unspoken assumptions and ordinary practices of conventional humanities research? Underlying the HuNI initiative is the recognition that cultural data is not economically, culturally, or socially insular, and in order to explore its dimensions fully, researchers need to collaborate across disciplines, institutions, and social locations. If we understand humanities research problems as comprising interdependent networks of institutional, social, and commercial practices, then new kinds of 'evidence', and new ways of organizing, accessing, and presenting this evidence, are critical for our enquiries.

To this end, HuNI provides a number of online research capabilities for humanities researchers to discover and work with the large-scale aggregation of data derived from different research domains and initially developed to solve different research questions. These capabilities enable researchers to create, save, and publish selections of data from HuNI; to analyse and manipulate the data; and to

export the data for reuse in external environments. The tasks that users can carry out in HuNI include searching and browsing the aggregated data, constructing private or shared virtual collections of HuNI data, exporting virtual collections for external analysis, and enhancing external databases for publishing into HuNI.

One of the most interesting and innovative features of HuNI, however, is the way it enables researchers to link entities within the HuNI aggregate through socially curated assertions. Using this feature, researchers are able to make statements about relationships between entities represented in the aggregated data. If, for example, they search the data aggregate and identify two entities in their result set that are related in some way, they can add a link between the two records and describe the nature of the relationship. The linking statement may be drawn from a suggested vocabulary of relationships, or the researcher may simply use free text. This feature also allows a researcher to assert that two entities are not related, in recognition that this kind of statement is also a key characteristic of humanities research.

To help visualize these social links, each entity has its own network graph, showing up to six degrees of separation, resulting in an expanding network of dynamic connections. These 'social linking' assertions are visible in the HuNI data aggregate. They may also appear in virtual collections assembled and published by individual users of the HuNI Virtual Laboratory. In this way, the 'social linking' of data forms the basis for researchers to create as well as browse network graphs within the HuNI Lab.

Crucially, the provenance of all these 'social linking' statements is also captured, enabling subsequent researchers to see who made each assertion. HuNI users can annotate these socially produced links with their own comments and assessments. Additionally, HuNI is an aggregate with a relationship to its own history. Researchers can trace how records (including the links between them) have changed over time. This capability recognizes that humanities research involves the mutual study of conceptual and temporal relationships. In other words, our research not only involves making connections between entities; it also involves assessing variations in cultural flow and network relationships through time. Each HuNI record is time-stamped, meaning that although researchers will always see the current view of a record, alongside its related records and assertions, they will also have the option to view how the record has changed since it was first harvested. The provenance information for each record, together with any curated assertions, is captured so that researchers can see when the records were harvested and by whom. A link to the originating data record at source is also provided in the user interface.

The capacity for HuNI users to assert data relationships in their own terms, in what we might call 'vernacular linking', is a central feature of the HuNI virtual laboratory. Instead of relying on a predetermined mapping to a detailed ontology, we are relying on researchers and community users to establish most of the connections within the heterogeneous data aggregate. This enables HuNI to capture the different disciplinary perspectives of users, rather than trying to fit all the data into a single normative framework. It also acknowledges the productive

differences that both define and link specific domains through a form of generative knowledge transfer. The opportunity to socially link data allows HuNI to encourage its users to share their knowledge and research findings in the form of specific assertions, and to discuss or debate these statements with each other.

This technology was developed in consideration of the need to balance the disciplinary imperatives for specific vocabularies and data structures with designing a service that is explicitly intended to transcend disciplinary boundaries and link related data effectively and meaningfully. Socially linked data as proposed by HuNI questions the use of standardized ontologies across a large-scale aggregation of heterogeneous humanities data. It suggests that 'linked open data' might equally be explored through 'open linking' as much as the more conventional emphasis on 'open data'. Paul Walk (2009) has very usefully summarized the various permutations of 'linked open data' as

- Data can be *open*, while **not** being *linked*
- Data can be *linked*, while **not** being *open*
- Data which is both *open and linked* is increasingly viable
- The *Semantic Web* can only function with data which is both *open and linked* [emphases in original]

To this summary HuNI might add,

- Linking can be open, while data is not.
- Linking and data can both be open.
- Linking that is both open and treated as data (time-stamped and retrievable) is increasingly viable.
- The Semantic Web is more meaningful with frameworks in which both data and linking is open.

The ontological frameworks that sit behind linked open data typically exist in *a priori* structures and are open to the kinds of criticisms that are frequently directed at humanities information and knowledge management strategies. Humanities data is heterogeneous, complex, inconsistent, interpretive, and frequently qualified. Ontologies 'make sense' of what might appear like flux through processes of agreed normalization and simplification. Ontologies are required for automated reasoning and typically rely on agreed common vocabularies and shared understandings of the semantic and conceptual structure of a domain. As Burrows and Verhoeven have noted (2014), in this context ontologies are limited in their ability to grapple with:

- *Variations in terminology*. The same concept or phenomenon may be described using different terms by different researchers, let alone by the wider community.
- *Vagueness of terminology*. The same term may be capable of referring to different concepts or phenomena, depending on the context.

- *Historical change.* The understanding of a knowledge domain is likely to have changed dramatically over time, along with the vocabulary and values used. Ontologies do not usually have a temporal dimension.
- *Multilingualism.* Much humanities research involves languages other than English, either for the subject of the research or for the research discourse itself. Most ontologies, on the other hand, use a standardized (and often quite formal) version of the English language.
- *Interdisciplinarity.* There is an obvious tension between (and even within) humanities disciplines in their terminology, methodologies, and intellectual models.

HuNI asks, why can't we have 'fluid ontologies' (Srinivasan and Huang 2005) that are not predefined but emergent and adaptive? What would it look like to enable ontologies to be co-created 'from below', to give ontologies a vernacular dimension? Can links themselves (and not just the data they connect) be more open? Do user-generated ontologies also open new lines of enquiry about how different communities make sense of the world?

This proposition is different from the vision painted by Clay Shirky in his advocacy for 'distributed classification'. An outspoken critic of both institutionally regulated data collection and its publication in formats to which members of the public have very little access, Shirky instead encourages information strategies such as collaborative tagging, folksonomies, distributed classification, or ethno-classification so that information users are enabled to create and aggregate their own metadata. But for Shirky these strategies are oriented to the same end, a (differently produced) version of information harmonization. For example, in his essay 'Ontology Is Overrated' (2005), Shirky argues that 'flipping' ontologies from a 'top-down' view of the world to one that is generated from the 'bottom up' will ultimately produce a new and more valid form of consensus rather than HuNI's recognition of multiple and sometimes discordant information stakeholders.

Vernacular ontologies such as that enabled by HuNI, on the other hand, incorporate a 'data ethics' that realizes that the quest for increasing standardization and alignment can be an intrusion, at the same time acknowledging that ontological relativism is equally distracting. Vernacular ontologies perform ethically insofar as the orthodox information management quest for coordination is replaced by cooperation; they accommodate instead the multiplication of difference and the social production-through-linking of other realities and experiences of the world. Vernacular ontologies champion the potential of open, participatory and collaborative linked data practices to produce new possibilities, for both knowledge and for linked data itself. Socially linked data opens the researcher to the associative. It asks, for instance, is it possible to organize information along expressive or ephemeral coordinates?

HuNI's 'social linking' brings to the fore the prosociality of databases—not only as artefacts of human action, imagination, ambition, and accomplishment (and failure), but in terms of bringing into prominence the ethical implications and possibilities of databases. Recent years have seen the rapid development of social

networking technologies (Baym and boyd 2012). While considerable investigation of the sociological and communication aspects of social networking has taken place, less attention has been paid to the potential for social networks to catalyse and enable humanities research itself—and even less to the larger question of what scholarship itself might mean in a digital ecosystem where sociality (rather than traditional systems for assessing academic merit) affords research opportunity and success. An implicit outcome of HuNI’s focus on socially linked data is that it also draws attention to our own relationships in research, and not only amongst our scholarly selves; it recognizes that a genuine commitment to open linking reveals and recasts the relationship between researchers and the public in terms that are larger than the conventional practices of ‘crowdsourcing’ content.

Case study 2

Songification: Enhancing Opportunities for Multimodal Knowledge

‘Songification’ is a method for enhancing auditory data that arose from work with The Ultimate Gig Guide (TUGG) database application (<http://tugg.me>). TUGG maps the Australian music industry at the level of individual performances from the mid-1960s. TUGG was specifically developed to better understand the flow of live music culture, through revealing the itineraries of bands, the socio-spatial location of music performances, and the various factors that have a role in the sustainability of music venues.

To capture band itineraries, the TUGG application represents a band’s gigs spatially on a Google map (what we like to call a form of ‘gigography’). But the scale of travel between venues is difficult to view in these maps—bands might play a sequence of gigs in the inner city and then move far afield to a country venue. For example, it is almost impossible to see the intricacies of adjacent movements around the city of Melbourne in the same map as a gig in the rural location of Colac (about 150 kilometres away) without losing a great deal of the detail. And because the Google maps are static, we cannot see the sequential order of a band’s gigs either.

To better represent these fluctuations of spatial scale and temporal sequence—and to ‘repatriate’ our research more meaningfully to the music community itself—we decided to try ‘sonifying’ our data (Hermann 2011). In particular, we were motivated by the idea of ‘thinking through’ our research in formats that make the most sense to the communities and industries we were studying. We didn’t want our research to be released as a fully formed, *a posteriori* afterthought to other academics only—any more than we want to just ‘deliver’ our results to the musicians and fans we think will be interested in TUGG. Avoiding this typical division between academics as agents (generating analysis) and nonacademic communities as the objects of research (generating content) meant ensuring that our research is undertaken in multimodal ways—acknowledging that although academics might enjoy written texts for developing and communicating their thinking, other communities might prefer to think visually or aurally.

The process of sonification was fairly straightforward and involved calculating the relative distance of gig venues from a central location, translating these figures into frequencies, transposing these into a frequency range that was recognizable to the human ear, and then translating these pro-rated frequencies into the nearest note in a C-major scale. The inharmonious result, however, defeated the idea that the sonified data could be easily shared. We then decided to take steps to enhance the sonifications, to enable clearer pattern detection and to honour the musical provenance of the data.

Songification entails an elaboration of the method by which sonification realizes data in the form of auditory values. By further transforming our research data into music, we created an improved experience for discerning the relations and rhythms in the data, demonstrating how enhanced auditory data design provides both a medium for aural intuition and an inclusive, ‘vernacular’ opportunity for nonprofessional research participation. It was also an opportunity to create some unique music by literally playing (with) the data.

We wrote backing tracks in the style of the bands we were studying and then added the itinerary data sequence as a lead guitar riff. Each gig/note was played in the succession in which it was originally performed—by date. The length or duration of each note was set to the number of days between one gig and the next gig. The longer the delay between gigs, the longer the note. The culmination of this process was a live performance by the TUGG team (all of whom play an instrument) of live performance data at the eResearch Australasia conference in 2013 (see Verhoeven 2013).

The songification method was specifically developed with an expanded view of what might constitute ‘open data’. In this context, exposing and enabling data sharing and reuse meant more than simply providing opportunities to make the data exportable into other data formats. As is typical in many data-driven projects, the TUGG research outcomes were available for public feedback online, but were presented in complex graphs and tables and were consequently inaccessible to the communities under study. Instead we were intent on re-presenting the data in forms and formats that made the most sense to the communities it described—in this case, the music industry.

The unprecedented opportunities presented by cultural data projects like TUGG, then, are that they lead not only to innovative methods for studying and understanding the creative industries and creative labour, but conversely, they enable us to simultaneously understand the creative potential of digital humanities research itself. And it changes the ‘ordering of things’ in other ways too, requiring a change of temporality—knowledge is always in process—not *a priori* nor *a posteriori*.

The specific point of a methodology like songification is to enable the live music community to engage with and contribute to our analysis as we develop it. We want them to be involved in helping us along with our analysis, to be much more than the passive recipients of our thinking. Songification is the instantiation of a belief that research itself is always in beta mode—that what is important is not the end result of a purely scholarly exercise but that the research that truly matters is inevitably iterative, multimodal, recursive, and co-created.

The TUGG researchers, for instance, developed diverse modes of participation and roles in the project that involved working with and through different media and information technologies—as researchers, performers, creators, audiences, and community members. From the outset, the TUGG research team included information managers, music industry analysts, IT developers, and geospatial scientists. We moved from one (or more) modes of action to another, depending on the aim of (and our role in) different activities (information manager/bass player/analyst). This form of modular research practice exceeds Biagioli's (2009) description of specialized interdisciplinary collaborations occurring in a limited temporal window and instead expands on our own capacity to simultaneously undertake multimodal roles. We were (non)academics, concurrently both the subjects and objects of our research undertaking, in chorus with the live music communities whose data we were studying/performing.

Songification explicitly recognizes the role of the media in shaping research, not just communicating it. The very act of playing the data as music had unexpected repercussions for our relationship to the research field—throwing even more significance onto the collaborative nature of large data-driven projects, emphasizing our ability to (really) listen, providing a (creative) medium for drawing together disparate ideas, and exercising our facility to focus on the present and the future simultaneously, to name a few. But even more curiously, playing the data/music has produced an affective response for some of our audiences, an unexpected feature of the project that remains to be further explored.

The development of songification was intended to be more than just an exercise in 'epistemic egalitarianism'. It opened new avenues for information and participation that also created opportunities for diversification of data analysis and associated research practices. The project began by recognizing that media producers and media scholars could both traverse the same pathways for discovery, development, discussion, and dissemination. What it had not anticipated was how productively the roles of media producer and scholar might themselves be entwined. The project also proposes further questions. How then might creative industry studies itself be rethought in a networked world where personal and professional identities so intimately overlap?

Conclusion

I wanted so much to write the story of the ocean. But what and where was the structure? I was, as they say, *all at sea*.

Winchester 2010, 24

The two case studies outlined in this chapter are interdisciplinary projects that were in part intended to elicit a reflection on the broader role of academics in society, which itself should not be seen as homogeneous but as composed of diverse (yet interlinked) interests and viewpoints. They were founded in the belief that there is value in reducing insularity in cultural and creative arts research—that an arm's-length critical distance is not always the most reliable way to grasp

something. Their operative view of openness and engagement is not restricted to the distribution of information but importantly also incorporates the formation of knowledge. And they recognize the value of the disaggregated cultural analysis that can be found outside the academy in places ranging from multinational corporations such as Google to the informal genealogies produced at the local library. In so doing, they aim to create new practices and politics of knowledge through digital scholarship—not just at the level of collection but at the level of analysis.

In the socio-hydrologic description of data inundation, that some digitally adept scholars might see themselves as latter-day Noahs is not surprising—building defensive arks to safeguard select and unique cultural collections, for example. Buried in here is a subsurface discourse of the ‘Digital Humanities as Noachian’, populated by ‘saviour academics’ returning the humanities to a time of prelapsarian order and structure. But data inundation can also have the effect of lifting our research imagination beyond familiar workflows into unfathomed waters. Here the digital humanities charts potential opportunities for rethinking our disciplines and for adjusting the course of our conduct in order to realize the challenge that databases present for the authority, influence, and purpose of the academic humanities.

Both HuNI and songification propose the digital creation of a social research ecology, made up of overlapping technologies, places, organizations, agents, concepts, and workflows. The effect of these intricate entanglements is to extend the scope of our scholarship at the same time as decentring it. For example, an emphasis on the ways that digital research produces new relations (and therefore also new forms of relationality) has a direct bearing on the production of our own digital scholarly selves. As David Beer and Mark Taylor (2013) suggest, engaged digital research initiatives ultimately mean

looking beyond the academy to see how ‘ordinary people’ are playing with the data themselves, how interns of companies are using it to understand and visualise their ‘customers’, to see what computer scientists and the like are producing with the data, to see what artists and others are creating with it.

In other words, as truly engaged academics, we don’t just learn from the data itself but also from the way that data is used and reused. An academic humility is at the heart of engagement-led digital scholarship, an acknowledgement that the ontological primacy of our own academic agency has (always) been dispersed and that the slipstreams of co-production cast their own currents.

Whilst this essay is explicitly about databases and the economies of engagement they afford—how they inflect what we do as digital scholars, how we do it, and whom we do it with—it is less obviously an essay about our thinking. Working with databases has repercussions for the course of our thoughts, appending to the intended use of the technologies in our research a recognition of the range of meanings made possible by these same technologies. As Sam Yazzie suggested in the example at the beginning of this paper, we might take pause to include the

sheep as part of our thinking, to think of ‘sheep’ (or even like a sheep) without succumbing to the recommendations of metaphor. By acknowledging nonhuman frameworks of what is meaningful, we may come close to Yazzie’s aspiration for ‘doing the sheep good’.

The case studies outlined in this chapter illustrate the benefits of an expanded and inclusive view of digital humanities and creative arts research; in which computation and communication, method and media, in combination enable us to explore the larger question of how we can work with technologies to co-produce, represent, analyze, convey and exchange knowledge. They suggest this exchange in the broadest sense—beyond the domestic diversions of academic interdisciplinarity. The key issue confronting the digital humanities is not the (largely self-defined) differences between academics but the perceived difference between academics as a whole and the community (both of which include the nonhuman). These case studies are a timely reflection on how our practices have lent weight to these perceptions and how an expanded and inclusive digital humanities and creative arts effort might redress them.

Acknowledgements

I would like to acknowledge the contributions of a number of people who have played a role in aspects of this paper: Thanks are due to the two anonymous reviewers whose acute assessments lent welcome improvements. The concept of ‘socially linked data’ was developed by the HuNI team, in particular the technical lead, Marco la Rosa, in conjunction with the HuNI Project Director (Deb Verhoeven); HuNI project manager, Alex Hawker; the HuNI semantic lead, Anne Cregan; and the HuNI product owner, Toby Burrows. Songification was elaborated by the author in collaboration with Alwyn Davidson, James Verhoeven, Alex Gionfriddo, and Peter Gravestock. Sound and video files relating to Songification can be found at <http://kinomatics.com/songification/>.

References

- Azuma, Hiroki. (2009). *Otaku*. Minneapolis: University of Minnesota Press.
- Baym, Nancy K., and danah boyd. (2012). ‘Socially Mediated Publicness: An Introduction’. *Journal of Broadcasting and Electronic Media* 56, no. 3 (September): 320–29.
- Beer, David. (2012). ‘Using Social Media Data Aggregators to do Social Research’. *Sociological Research Online* 17, no. 3 (August). Available at: <http://www.socresonline.org.uk/17/3/10.html>.
- Beer, David, and Mark Taylor. (2013). ‘The Hidden Dimensions of the Musical Field and the Potential of the New Social Data’. *Sociological Research Online* 18, no. 2 (May). Available at: <http://www.socresonline.org.uk/18/2/14.html>.
- Biagioli, Mario. (2009). ‘Postdisciplinary Liaisons: Science Studies and the Humanities’. *Critical Inquiry* 35, no. 4 (Summer): 816–35.
- Bryson, John. (2004). ‘What to Do When Stakeholders Matter’. *Public Management Review* 6, no. 1: 21–53.
- Burrows, Toby, and Deb Verhoeven. (2014). ‘Deploying Ontologies in the Humanities and Creative Arts’. Presentation at Expanding Horizons: Digital Humanities Australasia, University of Western Australia, 19 March.

- Derrida, Jacques. (1980). *Writing and Difference*. Chicago: University of Chicago Press.
- Di Baldassarre, G., A. Viglione, G. Carr, L. Kuil, J. L. Salinas, and G. Blöschl. (2013). 'Socio-Hydrology: Conceptualising Human-Flood Interactions'. *Hydrology and Earth Systems Sciences*, no. 17: 3295–303. DOI:10.5194/hess-17-3295-2013.
- Ellis, Carolyn. (2007). 'Telling Secrets, Revealing Lives: Relational Ethics in Research with Intimate Others'. *Qualitative Inquiry* 13, no. 1 (January): 3–29.
- Folsom, ed. (2007). 'Database as Genre: The Epic Transformation of Archives'. *PMLA* 122, (October): 1571–79.
- Fuller, Matthew, ed. (2008). *Software Studies: A Lexicon*. London: MIT Press.
- Gleick, James. (2011). *The Information: A History, a Theory, a Flood*. New York: Pantheon Books.
- Hayles, N. Katherine. (2012). *How We Think: Digital Media and Contemporary Technogenesis*. Chicago: University of Chicago Press.
- Hermann, Thomas, Andy Hunt, and John G. Neuhoff, eds. (2011). *The Sonification Handbook*. Berlin: Logos Publishing House.
- IDC. (2012). *Digital Universe Study*, December. Available at: <http://www.emc.com>.
- Jones, Steven E., and George K. Thiruvathukal. (2012). *Codename Revolution: The Nintendo Wii Platform*. Cambridge, MA: MIT Press.
- Kitchin, Rob, and Martin Dodge. (2011). *Code/Space: Software and Everyday Life*. Cambridge, MA: MIT Press.
- Liu, Alan. (2008). *Local Transcendence: Essays on Postmodern Historicism and the Database*. Chicago: University of Chicago Press.
- Mackenzie, Adrian. (2006). *Cutting Code: Software and Sociality*. Oxford: Peter Lang.
- Maher, Jimmy. (2012). *The Future Was Here: Commodore Amiga*. Cambridge, MA: MIT Press.
- Manovich, Lev. (2001). *The Language of New Media*. Cambridge, MA: MIT Press.
- Montfort, Nick, and Ian Bogost. (2009). *Racing the Beam: The Atari Video Computer System*. Cambridge, MA: MIT Press.
- National Oceanic and Atmospheric Administration (NOAA). (2014). 'About 96 Percent of Earth's Water Is in the Ocean'. 23 January. Available at: <http://oceanservice.noaa.gov>.
- Osborne, Thomas. (2004). 'On Mediators: Intellectuals and the Ideas Trade in the Knowledge Society'. *Economy and Society* 33 (2004): 430–47.
- Price, Kenneth M. (2009). 'Edition, Project, Database, Archive, Thematic Research Collection: What's in a Name?' *Digital Humanities Quarterly* 3, no. 3. Available at: <http://www.digitalhumanities.org>.
- Savage, Mike. (2009). 'Against Epochalism: An Analysis of Conceptions of Change in British Sociology'. *Cultural Sociology* 3, no. 2 (July): 217–38.
- Shirky, Clay. (2005). 'Ontology Is Overrated: Categories, Links, and Tags'. Available at: <http://www.shirky.com>.
- Srinivasan, Ramesh, and Jeffrey Huang. (2005). 'Fluid Ontologies for Digital Museums'. *International Journal on Digital Libraries* 5, no. 3 (May): 193–204.
- Verhoeven, Deb. (2006). *Sheep and the Australian Cinema*. Melbourne: Melbourne University Press.
- Verhoeven, Deb. (2013). 'Turning Gigs into Gigabytes (and Back Again): Songification and Live Music Data'. Available at: <http://kinomatics.com>.
- Walk, Paul. (2009). 'Linked, Open, Semantic?' Available at: <http://www.paulwalk.net>.
- Weller, Martin. (2011). *The Digital Scholar: How Technology Is Transforming Scholarly Practice*. London: Bloomsbury Academic.
- Winchester, Simon. (2010). *Atlantic: A Vast Ocean of a Million Stories*. London: Harper Press.
- Worth, Sol, and John Adair. (1972). *Through Navajo Eyes: An Exploration in Film Communication and Anthropology*. Bloomington: Indiana University Press.