Konferenca
Jezikovne tehnologije in digitalna humanistika
Ljubljana, 2018

Conference on
Language Technologies & Digital Humanities
Ljubljana, 2018

Data Mining Workspace Sensors:
A New Approach to Anthropology

Ajda Pretnar,* Dan Podjed†

*Laboratory of Bioinformatics
Faculty of Computer and Information Science
University of Ljubljana
Večna pot 113, SI-1000 Ljubljana
ajda.pretnar@fri.uni-lj.si

†Institute of Slovenian Ethnology
Research Centre of the Slovenian Academy of Sciences and Arts
Novi trg 2, SI-1000 Ljubljana
dan.podjed@zrc-sazu.si

Abstract

While social sciences and humanities are rapidly including computational methods in their research, anthropology seems to be lagging behind. However, this does not have to be the case. Anthropology is able to merge quantitative and qualitative methods successfully, especially when traversing between the two. In the following contribution, we propose a new methodological approach and describe how to engage quantitative methods and data analysis to support ethnographic research. We showcase this methodology with the analysis of sensor data from a University of Ljubljana's faculty building, where we observed human practices and behaviours of employees during working hours and analysed how they interact with the environment. We applied the proposed circular mixed methods approach that combines data analysis (quantitative approach) with ethnography (qualitative approach) on an example of a smart building and empirically identified the main benefits of our methodology.

1. Introduction

Social sciences and humanities are rapidly adopting computational approaches and software tools, resulting in an emerging field of digital humanities (Klein and Gold, 2016). Among these is anthropology, which is particularly suitable for traversing between quantitative and qualitative methods. Anthropologists study and analyse human behaviours and cultures, with a particular focus on long-term fieldwork as a methodological cornerstone of the discipline. With an increasing availability of data coming from social networks and wearable devices among other sources (Miller et al., 2016; Gershenfeld and Vasseur, 2014), anthropologists can easier than ever dive into data analysis and study humans and their societies, subcultures and cultures quantitatively as well as qualitatively.

With this contribution, we tentatively place anthropology in the field of digital humanities, mostly because the suggested approach is multidisciplinary and by analogy similar to the shifts between distant and close reading (Jänicke et al., 2015) in literary studies. Just like distant reading can offer an abstract (over)view of the corpus, quantitative analyses can give a researcher a broad understanding of the population she is investigating. And just like distant reading needs close reading to understand the style, themes, and subtle meanings of a literary work, so does data analysis need an ethnographic approach to contextualize the information and extract subtle meanings of individual human experience.

As Pink et al. (2017) suggest, there is value in investigating everyday data that reveal what is ordinary, what extraordinary and how to contextualize the two. In this contribution we expand the idea by employing the so-called *circular mixed methods approach* that combines qualitative research from anthropology and quantitative analysis from data mining. We consider mixed methods (Creswell and Clark, 2007; Teddlie and Tashakkori, 2009) as an integrative research that merges data collection, methods of research and philosophical issues from both quantitative and qualitative research paradigms into a singular framework (Johnson et al., 2007). We also stress the need for a circular research design, where we intentionally traverse between methods to continually verify and enhance our knowledge of the field. Circularity gives research flexibility and enables shifting perspectives in response to new information.

Our research began in October 2017 and currently monitors 14 offices at one of the University of Ljubljana's faculty buildings. We retrieved measurements from approximately 20 sensors from the SCADA monitoring system for the year 2016 and extrapolated behavioural patterns for different rooms and, more generally, room types through data visualization and exploratory analysis. The analysis showed specific patterns emerging in several rooms - there were some definite outliers in terms of working hours and room interaction.

We used computational methods to gauge new perspectives on human behaviour and invoke potentially interesting hypotheses. Data analysis provided several distinct patterns of behaviour and defined the baseline for workspace use. However, this approach was unable to provide us with a context for the data. Quantitative methods can easily answer the 'what', 'where' and 'when' type of questions, but

Konferenca
Jezikovne tehnologije in digitalna humanistika
Ljubljana, 2018

Conference on
Language Technologies & Digital Humanities
Ljubljana, 2018

struggle with the 'why'. At that stage, we employed long-term fieldwork and ethnography as the main methods of anthropology. We conducted interviews with room occupants to explain what the uncovered patterns mean and why people behave the way they do.

The main goal of our study was to demonstrate how anthropologists can use statistics and data visualisation to establish the essential facts of the observed phenomena and how the traditional anthropological methods, which have not significantly changed since the early 20th century (Malinowski, 1922), can be complemented and upgraded by data analysis. We call this a circular mixed methods approach, where circular implies continual traversing between qualitative and quantitative methods, between fieldwork and data analysis. The present contribution applies the proposed methodology to sensor data obtained from a smart building and with a combination of data mining and ethnographic fieldwork establishes both a wide and deep understanding of human behaviour in a workplace setting.

2. Related Work

While digital humanities became a full-fledged field in the last couple of decades (Hockey, 2004), anthropology seems to be left of out its spectrum. Some authors suggest anthropology would be more concerned with digital as an object of analysis rather than as a tool (Svensson, 2010). However, there have been several attempts to include computational methods and quantitative analyses into anthropological research. Already in the 1960s, anthropologists looked at using computers for organisation of anthropological data and field notes (Kuzara et al., 1966; Podolefsky and McCarty, 1983). Progress in text analysis, coding facts, and comparative studies in linguistics (Dobbert et al., 1984; White and Truex, 1988) followed suit.

However, only lately there has been some digital shift in the discipline. Digital anthropology turned disciplinary attention to the analysis of online worlds, virtual identities, and human relationships with technology. For example, Bell (2006) gave a cultural interpretation of the use of ICTs in South and Southeast Asia, Nardi (2010) explored gaming behaviour of the World of Warcraft, Boellstorff (2015) investigated alternate online worlds of the Second Life, and Bonilla and Rosa (2015) described how to use hashtags for ethnographic research. Moreover, a discussion has been opened on what does 'big data' mean for social sciences and how to ethically address its retrieval and analysis (Boyd and Crawford, 2012; Mittelstadt et al., 2016).

There was a discussion on the methodological front as well. Anderson et al. (2009) argue for a method that combines the ethos of ethnography with database mining techniques, something the authors call 'ethno-mining'. Similarly, Blok and Pedersen (2014) look at the intersection of 'big' and 'small' data to produce 'thick' data and include research subjects as co-producers of knowledge about themselves. Finally, Krieg et al. (2017) not only elaborate on the usefulness of algorithms for ethnographic fieldwork, but also show in detail how to conduct such research in an example of online reports of drug experiences.

3. Anthropology vs. Data Analysis

For an anthropologist, statistical and computational analysis is not the first thing that comes to mind when developing research design and methodology. Anthropologists are trained to observe phenomena in the field, talk to people, spend time with them, participate in daily activities, and immerse themselves in topics of interest (Kawulich, 2005; Marcus, 2007). This type of information gives us detailed stories of human lives, uncovers meanings behind rituals, habits, languages, and relationships, and provides a coherent explanation of the researched phenomena. So why would anthropologists even have to include data analysis in their studies? Why and when is such an approach relevant?

Sometimes, the phenomena that anthropologists are trying to explain occur in different places at the same time and are impossible to observe simultaneously. It could be that anthropologists know little of the topic they are exploring and have yet to generate their research questions. Or the nature of the phenomenon lends itself nicely to computational analysis. For example, behaviour of many individuals is difficult to observe in real time, especially if we want to observe them at once in different locations. Sensors, on the other hand, can track behaviours of these individuals independently (Patel et al., 2012) and therefore enable a detailed comparative analysis. With a large amount of measurements, researchers can also observe seasonal variations, similarity of users, and changes through time.

Data analysis also helps us define the parameters of our research field and establish what is an ordinary and what is an extraordinary behaviour. Visualisations in particular are excellent tools for exploring and understanding frequent patterns of behaviour and outliers. When done well, visualisations harness the perceptual abilities of humans to provide visual insights into data (Fayyad et al., 2002, p. 4). Moreover, they provide a new perspective on a phenomenon and help generate research questions and hypotheses. Once we know how our research participants behave (or communicate if we are observing textual documents or establish social ties if we are observing social networks), we can enter the field equipped with knowledge and information to verify and contextualise.

Finally, large data sets are particularly appropriate for computational analysis. While 'big data' became a popular buzzword in data science, anthropologists most likely will not be dealing with millions of data points that can be analysed only with graphics processing units (GPUs). However, even ten thousand observations are too much for a researcher to make sense of. For such data, we need software tools and visualisations, which provide an overview of the phenomenon, plot typical patterns, and enable exploring different sub-populations.

4. Data Ethics and Surveillance Technologies

Data ethnography inevitably raises questions of ethics, just like sensor data inevitably raise the question of surveillance. Both topics are too broad for the scope of this contribution but let us briefly touch upon them. Research ethics, in particular sensitivity to the potential harm a study

Konferenca
Jezikovne tehnologije in digitalna humanistika
Ljubljana, 2018

Conference on
Language Technologies & Digital Humanities
Ljubljana, 2018

could elicit, is one of the core questions of anthropology, which is so deeply immersed in the personal human experience. A solid deontological paradigm is crucial for working with not only sensitive data but any human-produced data. In this sense, we follow the principles of positivist ethics which call for human dignity, autonomy, protection, maximizing benefits and minimizing harm, respect, and justice (Markham et al., 2012; Halford, 2017).

As for surveillance, we propose a distinction between surveillance and monitoring. Surveillance implies guiding actions of surveilled subjects, while monitoring proposes a more passive stance of observing behaviour. The present study was not designed to guide behaviour but to observe and understand, hence being more monitoring than surveillance focused. And even if we consider it surveillance-like, Marx (2002) proposes "a broad comparative measure of surveillance slack which considers the extent to which a technology is applied, rather than the absolute amount of surveillance", meaning that the extent to which surveillance is harmful is the power it holds for the user. The case of sensor data of a smart building that monitors only neutral human behaviour [1], falls to the soft side of power, which, in the opinion of the authors, deserves some surveillance slack. Nevertheless, we strived to uphold high ethical standards for handling the data and disseminating the results, mostly by employing "ongoing consensual decision-making" (Ramos, 1989) by informing our participants of the purpose of the research, which data are being collected and how the findings are going to be presented.

5. Data Preprocessing

In our study, we have observed sensor measurements from a faculty which is considered to be a state-of-the-art smart building in Slovenia. Each room in the building is equipped with a temperature sensor and sensors on windows that track when they are open or closed. Doors have electronic key locks that track when the room is occupied. There were altogether 11 sensor measurements, with an additional 8 measurements coming from the weather station located on the building's rooftop. In-room sensor reports the room temperature, set temperature, ventilation speed, daily regime, and so on, while the weather station reports the external temperature, light, rainfall, etc. One of the most important measurements is the daily regime, which has four values, each representing a state of the overall room setting. When a person is present in the room, the regime is comfort (value = 0) and when a window is open, the regime is off (value = 4). If the room is vacant, the regime goes to night (value = 1) or standby (value = 3)[2].

We retrieved 55,456 recordings for 14 rooms of different types, namely 5 laboratories, 6 cabinets, and 3 administration rooms. Measurements are recorded bi-hourly and stored in the SCADA monitoring system. We decided to observe the year 2016 and later compare it to 2017. The results in the paper refer only to 2016. The rooms are

anonymised to ensure data privacy and results for two of the rooms are not reported at the request of their occupants.

We performed extensive data cleaning and preprocessing and removed data points with missing values (Table 1). We considered *daily regime* as our most important variable since it reports a presence in the room or the opening of windows. Concurrently, we removed data points where the daily regime was *comfort* throughout the day[3].

For the analysis, we retained only one feature, namely *daily regime*, since, as mentioned above, this was the feature that registered human behaviour the best. We also generated additional features, such as the day of the week and room type (cabinet, laboratory, and administration).

In the second part of the analysis, we created a transformed data set where we merged daily readings for a room into one 'daily behaviour' vector (Table 2). In the new data set, each room has a daily recording, where the new features are values of the daily regime at each hour. Since sensors only record the state every two hours, we filled missing values with the previous observed state. For example, if the original vector was $\{0, ?, 0, ?, 1, ?, 1\}$, we imputed missing values to get $\{0, 0, 0, 0, 1, 1, 1\}$. As we were interested only in the presence in the room, we put 0 where daily regime was 1 (night) or 3 (standby) and 1 where it was 0 (comfort) or 4 (window open), discarding the information on specific temperature regimes. This gave us the final daily behaviour vector which we could compare in time and between rooms.

6. Results

First, we wanted to see how rooms differ by room occupancy alone. We hypothesised there will be a significant difference in occupancy between laboratories and cabinets since the presence of more people in a space extends the occupancy hours (no complete overlap of working time).

We took the first data set with bi-hourly recordings and removed readings where the daily regime was either 1 (night) or 3 (standby) because these readings indicate the room was not occupied. Afterwards, we computed the contingency matrix of room occupancy by the day of the week, which shows how many times per year a room was occupied on a certain day. We visualised the result in a line plot (Figure 1). We can notice that laboratories have a higher presence on Saturday and Sunday than the other rooms.

Moreover, N and O are the top two rooms by occupancy. We know that these two rooms belong to a single laboratory and are separated with a permanently open door. These two rooms are occupied by the largest number of people and since the employees of the faculty have a somewhat flexible working time, the dispersion of working time is expectedly the highest in rooms with the most occupants (smallest overlap in working time among employees). N and O are also among the few rooms where occupancy goes up towards the end of the week.

F and B are also laboratories, both displaying similarly high presence across the week. On the bottom of the plot there are cabinets, namely G, K, F. Unsurprisingly, cabinets display lower occupancy rates than laboratories, since cabinets are used by a single person and hence no overlap is

---

[1]We consider neutral human behaviour a behaviour which does not explicitly convey sensitive information.

[2]Standby is activated on workdays as a transitory setting between night and comfort regime.

[3]We considered a constant comfort regime an error in the reading since a 24-hour workday is highly unlikely.

Konferenca
Jezikovne tehnologije in digitalna humanistika
Ljubljana, 2018

Conference on
Language Technologies & Digital Humanities
Ljubljana, 2018

Date	Room temp	Daily regime	Room
2016-01-01 02:10:00	20.94266	1	C
2016-01-01 02:10:00	21.65854	1	B
2016-01-01 02:10:00	20.63234	1	K
2016-01-01 02:10:00	22.41270	1	D
2016-01-01 02:10:00	20.25890	1	M
2016-01-01 07:10:00	21.45220	3	C

Table 1: Original data

Date	0am	1am	2am	3am	...	Room	Day	Type
2016-01-01	0	1	1	0	...	C	Fri	laboratory
2016-01-01	0	0	0	0	...	B	Fri	laboratory
2016-01-01	0	0	1	1	...	K	Fri	administration
2016-01-01	0	0	0	1	...	D	Fri	cabinet
2016-01-01	0	1	1	1	...	M	Fri	administration
2016-01-02	0	0	1	1	...	C	Sat	laboratory

Table 2: Data transformed into a behaviour vector. 1 denotes presence in the room, meaning daily regime value was either 0 (comfort) or 4 (window open).
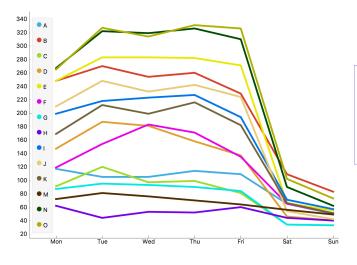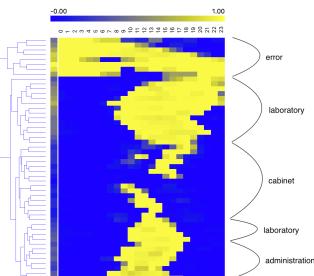

Figure 1: Occupancy of the rooms for each day of the week.

Figure 2: Occupancy by the hour of the day. Distinctive room-type patterns emerge.

possible. They are also functional rooms, used predominantly for meetings, office hours, and other intermittent work of professors.

With the second room occupancy data set, we made an analysis of behavioural patterns by the time of the day. We observed occupancy by room type in a heat map where 1 (yellow) means presence and 0 (blue) absence. Visualisation in Figure 2 is simplified by merging similar rows with k-means (k=50) and clustering by similarity (Euclidean distance, average linkage and optimal leaf ordering). Such simplification joins identical or highly similar patterns into one row and rearranges them so that similar rows are put closer together.

Clustering revealed that occupancy sequence highly depends on the room type. There were some error data, where sensors recorded presence at unusual hours (for example during the night consistently across all rooms). But despite some noise in our data, we can distinguish between typical laboratory, administration and cabinet behaviour, since our error data constitute a separate cluster (Dave, 1991). Cabinets again show the lowest occupancy with presence recorded sporadically across the day. Normally, university lecturers spend a large portion of their time in lecture rooms and in their respective laboratories. This is why occupancy of cabinets is so erratic and does not display a consistent pattern. Laboratory occupants, on the other hand, usually come late and stay late, while administration staff work regularly from 7:00 a.m. to 4:00 p.m. They both display fairly consistent behaviour.

We visualised the same data set in a line plot, which

Konferenca
Jezikovne tehnologije in digitalna humanistika
Ljubljana, 2018

Conference on
Language Technologies & Digital Humanities
Ljubljana, 2018

shows the frequency of attributes on a line. This way, we can better observe differences between individual rooms at each time of the day and where specific peaks (high frequencies) happen. Figure 3 displays the occupancy ratio at a specific time of the day, while Figure 4 shows the ratio of window opening [4]. Several interesting observations emerge. In both cases, room O is skewed to the right, meaning its occupants work at late hours and open windows while working. Conversely, room J is skewed to the right, indicating its occupants start work earlier than most. There is also a distinct peak in window opening at around lunch time.
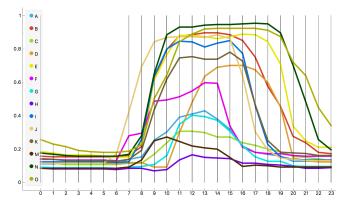

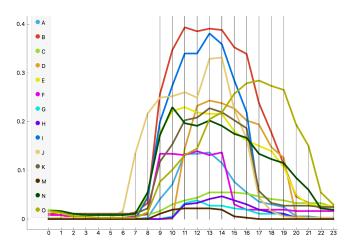Figure 3: Room occupancy by the time of the day.

Figure 4: Window opening frequency by the time of the day.

In most rooms, people are opening windows from late morning to early afternoon. Again, not surprising, considering this is their peak working time. This is a great indicator for an ethnographer if he or she wants to observe window interaction (who does it, is there a consensus on whether or not it should be opened, does this happen more frequently after lunch...). Looking at the data, the best time for observing the specified behaviour is between 10:00 a.m. and 1:00 p.m. Accordingly, data analysis can also serve as a guide for ethnographic field work.

---

[4] 1 would mean the room was always occupied and 0 that the room was never occupied at a specific time of the day.

7.   Ethnography Comes In

Data analysis revealed some interesting patterns in the use of working spaces:

- laboratories work more on weekends,
- rooms N and O work late,
- room J starts the day early and opens windows at lunchtime, and
- in rooms H, N and O the occupancy goes up towards the end of the week

How can we explain this? While the data gave us clues, the answers lie with the people. Substantiating analytical findings with fieldwork ethnography is crucial for understanding the data. We conducted semi-structured interviews with the rooms' occupants to discover what those patterns mean and why a certain behaviour occurs.

Laboratories have a higher weekend occupancy since they offer a quiet place to work for PhD students who are either catching deadlines for publishing papers or using their 'off time' for some in-depth research. Room B, in particular, seems to like working at weekends and we were able to identify an individual who often comes to work on Saturdays. In the interview, he[5] told us this was time when he was able to really focus on his dissertation.

Rooms N and O are quite similar in terms of presence although room N displays a tendency to work the latest. By observing inhabitants in this room and talking to them, we identified an individual who preferred to work in the late afternoon and evening. Since, as mentioned above, working time is flexible at the studied faculty, he adjusted his working hours to suit his preferences. He also prefers fresh air to artificial ventilation and opens the windows whenever possible. This accounts for the skew to the right for room O in Figure 4.

The increased productivity in rooms N, O, and H towards the end of the week is explained by the fact that Fridays are working sprints for occupants of these three rooms. The case of room H is particularly interesting. This is the room with the overall lowest occupancy, yet the room is most frequented on Fridays, unlike in most other rooms, where the occupancy decreases towards the end of the week. Room H is the cabinet of a professor who runs laboratories N and O. He is also a part of the Friday working sprints, hence the peak. Yet he is very social and prefers to work in the laboratory with colleagues, rather than alone in the cabinet. This explains the overall low and erratic occupancy of his room during the rest of the week.

The skewed peak for room J in Figure 3 is again interesting. The occupant of this room admitted he prefers coming to work earlier to make the most of the day. He stressed several times that daylight is important to him and by shifting working time to earlier hours, he was able to leave early and use the rest of the day for himself. He also said he was the most productive in early mornings since these were the quietest parts of the day. Personal preferences evidently affected the discovered patterns of workday behaviour.

Such a circular methodological approach, where the researcher traverses between data analysis, observation, and

---

[5] Pronoun he is used for both males and females.

Konferenca
Jezikovne tehnologije in digitalna humanistika
Ljubljana, 2018

Conference on
Language Technologies & Digital Humanities
Ljubljana, 2018

ethnography, has several benefits. Large data collections can be effectively and rapidly analysed with computational means. Visualisations, moreover, substantiate the findings and enable researchers to uncover relations, patterns, and outliers in the data. Hence, data analysis can help generate hypotheses and questions for the research. This cuts down the time required to get familiar with the field. A researcher can come into the field equipped with potentially interesting hypotheses and test them almost immediately.

Looking at the data alone, however, we would be unable to determine what any of those patterns and outliers mean. To truly understand them, we need to immerse ourselves in the field, ask questions and observe how people behave and create their habits and practices. While quantitative analysis provides us with clues, qualitative approaches, such as ethnography and fieldwork, explain those clues and substantiate the superficial knowledge of the field acquired in the first research phase. Metaphorically speaking, data analysis is great for scratching the surface, while ethnography excels at digging deeper.

8. Conclusion

In this paper, we have shown the how to combine quantitative and qualitative methods for anthropological research. While the findings are still preliminary and based on a limited sample, they nevertheless pinpoint aspects of data analysis that benefit from ethnographic insight and vice versa.

With the increasing availability of data, especially from sensors, wearable devices, and social media, anthropologists can use computational methods and data analysis to uncover common patterns of human behaviour and pinpoint interesting outliers. Quantitative methods have proven useful when dealing with large data sets. In such cases, an analysis without digital tools is virtually impossible, while visualisations offer new insight into the problem and help present the data concisely. In addition, quantitative approaches also increase the reproducibility of research.

However, patterns emerging from such analysis can hardly ever be explained with data alone. We argue that data analysis can generate new hypotheses and research questions (Krieg et al., 2017) and provide a general overview of the topic. Conversely, ethnography substantiates analytical findings with the context and story behind the data. Going back and forth, from quantitative to qualitative and *vice versa*, enables researchers to establish a research problem as suggested by the data, gauge new perspectives on the known problems, and account for outliers and patterns in the data. Circular research design enhances the quality of information, which does not have to derive solely from a quantitative or qualitative approach. By combining the two, we are using a research loop that ensures both sets of data get an additional perspective - quantitative data are verified with ethnography in the field, while ethnographic data become supported with statistically relevant patterns.

Such methods are already, to a certain extent, employed in digital anthropology (Drazin, 2012), but they are gaining more prominence in mainstream anthropology as well (Krieg et al., 2017). By establishing a solid methodological framework for quantitative analyses in relation to qualitative ones, we do not only strengthen the subfield of computational anthropology, but also provide new perspectives and research ventures to anthropology and emphasise its relevance for studying lifestyles, habits, and practices in data-driven societies.

9. References

Ken Anderson, Dawn Nafus, Tye Rattenbury, and Ryan Aipperspach. 2009. Numbers have qualities too: Experiences with ethno-mining. In *Ethnographic Praxis in Industry Conference Proceedings*, volume 2009, pages 123–140. Wiley Online Library.

Genevieve Bell. 2006. Satu keluarga, satu komputer (one home, one computer): Cultural accounts of icts in south and southeast asia. *Design Issues*, 22(2):35–55.

Anders Blok and Morten Axel Pedersen. 2014. Complementary social science? quali-quantitative experiments in a big data world. *Big Data & Society*, 1(2):2053951714543908.

Tom Boellstorff. 2015. *Coming of Age in Second Life: An Anthropologist Explores the Virtually Human*. Princeton University Press.

Yarimar Bonilla and Jonathan Rosa. 2015. # ferguson: Digital protest, hashtag ethnography, and the racial politics of social media in the united states. *American Ethnologist*, 42(1):4–17.

Danah Boyd and Kate Crawford. 2012. Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication & Society*, 15(5):662–679.

John W Creswell and Vicki L Plano Clark. 2007. Designing and conducting mixed methods research.

Rajesh N Dave. 1991. Characterization and detection of noise in clustering. *Pattern Recognition Letters*, 12(11):657–664.

Marion Lundy Dobbert, Dennis P McGuire, James J Pearson, and Kenneth Clarkson Taylor. 1984. An application of dimensional analysis in cultural anthropology. *American Anthropologist*, 86(4):854–884.

Adam Drazin. 2012. Design anthropology: Working on, with and for digital technologies. *Digital Anthropology*, pages 245–65.

Usama M Fayyad, Andreas Wierse, and Georges G Grinstein. 2002. *Information Visualization in Data Mining and Knowledge Discovery*. Morgan Kaufmann.

Neil Gershenfeld and JP Vasseur. 2014. As objects go online; the promise (and pitfalls) of the internet of things. *Foreign Aff.*, 93:60.

Susan Halford. 2017. The ethical disruptions of social media data: Tales from the field. In *The Ethics of Online Research*, pages 13–25. Emerald Publishing Limited.

Susan Hockey. 2004. The history of humanities computing. *A Companion to Digital Humanities*, pages 3–19.

Stefan Jänicke, Greta Franzini, Muhammad Faisal Cheema, and Gerik Scheuermann. 2015. On close and distant reading in digital humanities: A survey and future challenges. In *Eurographics Conference on Visualization (EuroVis)-STARs. The Eurographics Association*.

Konferenca
Jezikovne tehnologije in digitalna humanistika
Ljubljana, 2018

Conference on
Language Technologies & Digital Humanities
Ljubljana, 2018

R Burke Johnson, Anthony J Onwuegbuzie, and Lisa A Turner. 2007. Toward a definition of mixed methods research. *Journal of Mixed Methods Research*, 1(2):112–133.

Barbara Kawulich. 2005. Participant observation as a data collection method. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, 6(2).

Lauren F Klein and Matthew Gold. 2016. Digital humanities: The expanded field. *Debates in the Digital Humanities*.

Lisa Jenny Krieg, Moritz Berning, and Anita Hardon. 2017. Anthropology with algorithms? *Medicine Anthropology Theory*, 4(3).

Richard S Kuzara, George R Mead, and Keith A Dixon. 1966. Seriation of anthropological data: A computer program for matrix-ordering. *American Anthropologist*, 68(6):1442–1455.

Bronislaw Malinowski. 1922. *Argonauts of the Western Pacific: An Account of Native Enterprise and Adventure in the Archipelagoes of Melanesian New Guinea*. G. Routledge & Sons.

George E Marcus. 2007. Ethnography two decades after writing culture: From the experimental to the baroque. *Anthropological Quarterly*, 80(4):1127–1145.

Annette Markham, Elizabeth Buchanan, AoIR Ethics Working Committee, et al. 2012. Ethical decision-making and internet research: Version 2.0. *Association of Internet Researchers*.

Gary T Marx. 2002. What's new about the" new surveillance"? classifying for change and continuity. *Surveillance & Society*, 1(1).

Daniel Miller, Elisabetta Costa, Nell Haynes, Tom McDonald, Razvan Nicolescu, Jolynna Sinanan, Juliano Spyer, Shriram Venkatraman, and Xinyuan Wang. 2016. *How the World Changed Social Media*. UCL press.

Brent Daniel Mittelstadt, Patrick Allo, Mariarosaria Taddeo, Sandra Wachter, and Luciano Floridi. 2016. The ethics of algorithms: Mapping the debate. *Big Data & Society*, 3(2):2053951716679679.

Bonnie Nardi. 2010. *My Life as a Night Elf Priest: An Anthropological Account of World of Warcraft*. University of Michigan Press.

Shyamal Patel, Hyung Park, Paolo Bonato, Leighton Chan, and Mary Rodgers. 2012. A review of wearable sensors and systems with application in rehabilitation. *Journal of Neuroengineering and Rehabilitation*, 9(1):21.

Sarah Pink, Shanti Sumartojo, Deborah Lupton, and Christine Heyes La Bond. 2017. Mundane data: The routines, contingencies and accomplishments of digital living. *Big Data & Society*, 4(1):2053951717700924.

Aaron Podolefsky and Christopher McCarty. 1983. Topical sorting: A technique for computer assisted qualitative data analysis. *American Anthropologist*, 85(4):886–890.

Mary Carol Ramos. 1989. Some ethical implications of qualitative research. *Research in Nursing & Health*, 12(1):57–63.

Patrik Svensson. 2010. The landscape of digital humanities. *Digital Humanities*.

Charles Teddlie and Abbas Tashakkori. 2009. *Foundations of Mixed Methods Research: Integrating Quantitative and Qualitative Approaches in the Social and Behavioral Sciences*. Sage.

Douglas R White and Gregory F Truex. 1988. Anthropology and computing: The challenges of the 1990s. *Social Science Computer Review*, 6(4):481–497.