Entering the Space of Data

Joanna Sleigh
joannasleigh@gmail.com

1 Introduction

> Technology is neither good nor bad; nor is it neutral… technology's interaction with the social ecology is such that technical developments frequently have environmental, social, and human consequences that go far beyond the immediate purposes of the technical devices and practices themselves. (Kranzberg 1986: 545).

We live in an increasingly interconnected globalized world (Unwin 2009: 15). It is the era of big data. A time, wherein every second, hundreds of thousands of activities and personal experiences are documented and publicly shared across installations of online networks. This topic is an exciting space for exploration, so much so, that a diversity of individuals are entering the field. This includes computer scientists, economists, mathematicians, sociologists, artists and scholars.

As with other socio-technical phenomena, there is both utopian and dystopian rhetoric surrounding big data. For data can be both a powerful tool that offers new insights into a diversity of areas, while at the same time it can invade privacy, collapse civil freedoms and increase state control. This leads to questions of representations and access. Data is a communally produced product, it is an invisible layer documenting our lives, however what is the experience of accessing, analysing and representing it?

This practice led study explores the patterns and processes of developing a lived-world algorithm for the abstraction of an interconnected world. With methodology based in visual anthropology, I seek to unpack what effect working with data and data visualizations has on the individual for whom working with big data in unfamiliar. First I will provide details of the case site, La Cura, followed by the theoretical framework and methodology for this research. This will be ensued by a synthesis of thematically analyzed data and recurring concepts that emerged during research. Finally, I propose a representational design for the visualization of this anthropological data.

Fig. 1: *Sleigh, Joanna, Workshop Room. 2016, Photograph. La Cura Villa Strozzi, Italy.*

Fig. 2: *Sleigh, Joanna, Court yard. 2016, Photograph. La Cura Villa Strozzi, Italy.*

2   Case Site – La Cura

We experience a massive state of interconnectedness…"
(Salvatore Iaconesi. 2016).

The La Cura Summer School took place between the 22nd-26th of August 2016 in Florence, Italy. Developed along a transdisciplinary path composed of theoretical dialogues, meetings, design sessions, creative conditions and operational development paths, the La Cura program centered on examining the transdisciplinary topic of what it means to inhabit the planet in the era of hyper-connection (http://la-

cura.it/summerschool). The data mining tool 'Human Ecosystems'[1] was designed and used specifically for this purpose, and was of integral value to the project's output of an interactive installation at design and art museum, La Triennale di Milano, in Milan.

The La Cura program was the initiative of the 'Near Future Design Laboratory' Nefula, a distributed Laboratory focused on the Near Future Design methodology, and was led by both Salvatore Iaconesi, an interaction designer, robotics engineer, artist and hacker, and his partner Oriana Persico, an artist, writer and researcher in participatory policies and digital inclusions. La Cura was comprised of a collective of artists, researchers, designers, architects, philosophers, engineers, physicists, computer scientists, legal scholars, poets, thinkers from around the world. While the majority of participants were from Italy, others also came from Russia, America, Australia and Germany. Below is a diagram made during the La Cura program that documents the individuals involved.

The physical site of La Cura was the Villa Strozzi, in Florence, and this was turned into a transdisciplinary laboratory of science, arts, technology and design. Integrating the disciplines of Cybernetics, Ecology, and the Theory of Systems and Complexity, the collective worked with big data, social networks, wearable technologies, info-aesthetics and had philosophical and theoretical involvement. This was made possible through two data analysis tools, Human Ecosystems[2] and Circumplex Model of Affect.[3]

---

[1] A tool, downloadable from Git.Hub that enables anybody to search and analyze data on social network systems.

[2] Human Ecosystems is a data mining tool that captures data in 29 languages from multiple social networks in the public sphere. This tool was created to enable anyone to gain better understandings of society, communities, cultures, and to create innovative services, to organise and to act. - https://github.com/xdxdVSxdxd/HumanEcosystems.

[3] The Circumplex Model of Affect was combined with Natural Language Analysis. Together these tools enabled an analysis of the emotions based on the language used. Some scholars argue that the Circumplex model is limited and stands in contrast to theories of basic emotions (Posner, Russell & Peterson, 2008). Image source: wangchenwei.com.
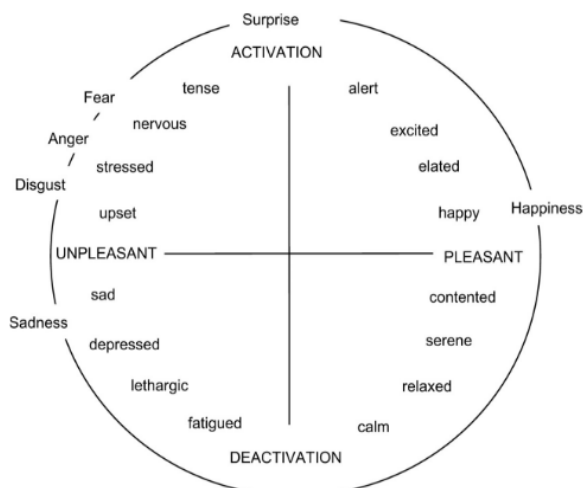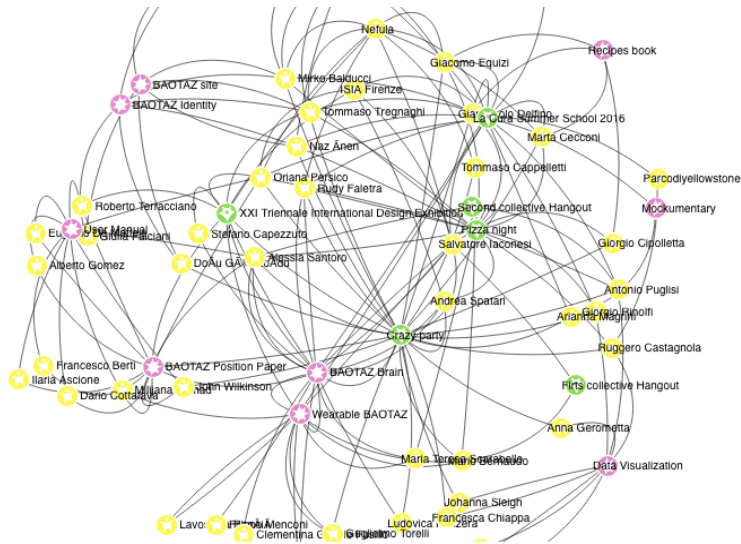
Fig. 3. *Baotaz Commons, Documentation of everyone invovled in the program. 2016, Screenshot. http://baotaz.net/ (accessed 9 November 2016).*

The output of this week long experience of La Cura was the establishment of collaborative platform 'Baotaz' (http://baotaz.net/) and an installation / wearable technology that was exhibited on the 3rd and 4th of September 2016, at the 21st Triennale di Milano International Exhibition, as an element of the 'Share Knowledge' conference. This installation let users experience "*the relationships and emotions generated by our interconnected lives*" (http://baotaz.net/). Specifically, it was a giant silicon brain, a headpiece, and a data visualization of social media representations of the word 'inhabiting' and 'un-inhabiting'. The data collected from the public realms of social media were transformed into electronic sensations that could be felt when wearing the connected headpiece.
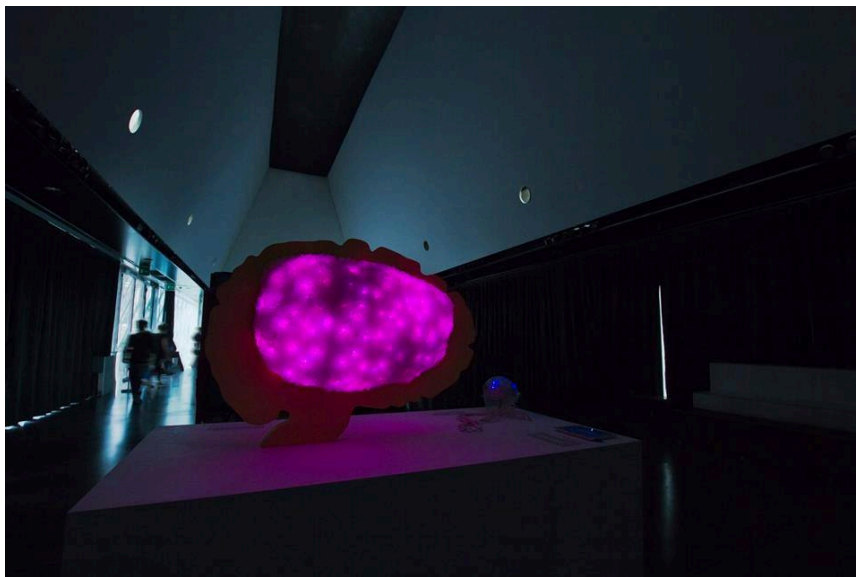

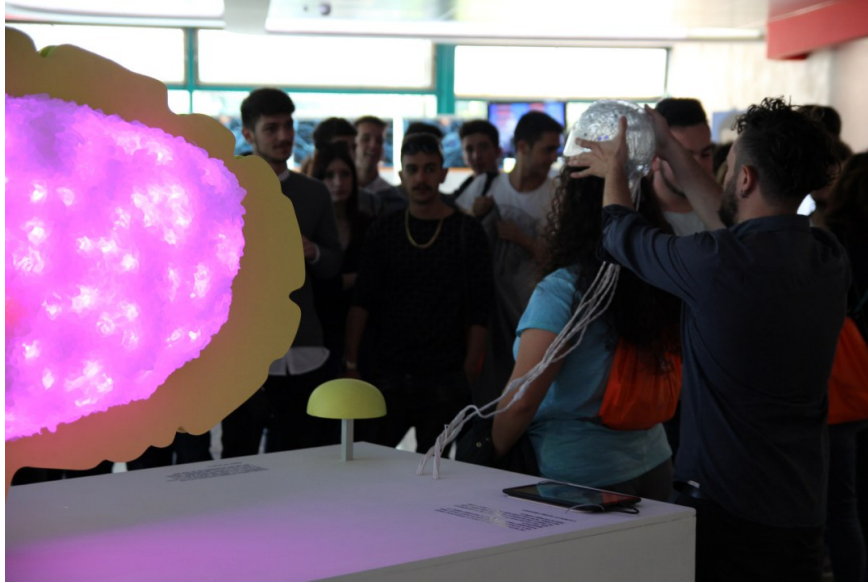Fig. 4. *Baotaz Commons, Boataz Brain sculpture. 2016, Screenshot. http://baotaz.net/ (accessed 9 November 2016).*

28

Fig. 5. *Baotaz Commons, Boataz Wearable device that tranmists collected data from Human Ecosystems directly into the body. 2016, Screenshot. http://baotaz.net/ (accessed 9 November 2016).*
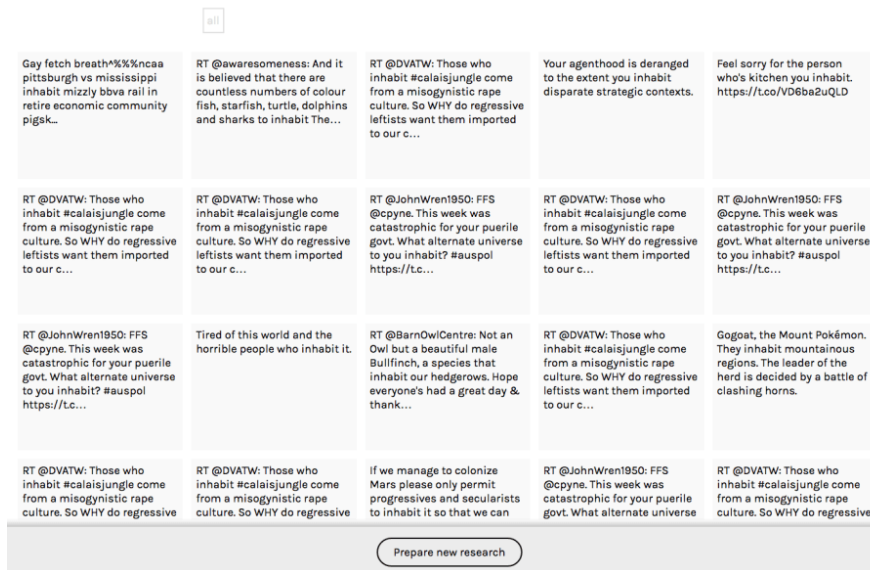

Fig. 6. *Baotaz Commons, Data Visualisation displaying the complexity of what it means to inhabit this planet. 2016, Screenshot. http://baotaz.net/ (accessed 9 November 2016).*

2 Methodology & Theoretical Framework

"Data has emerged as a system of knowledge that is already changing the objects of knowledge" (Boyd & Crawford 2012).

It is important to place this study within contemporary social thought. Firstly, because the very notion of what 'big data' refers to has changed as a socio-technical phenomenon (Boyd & Crawford 2012). Originally the term was used within the sciences to refer to data sets so large that computers were required (Manovich 2011). Today, the term has expanded. It now references not only large amounts of data, but also the tools and procedures used to manipulate and analyze data, along

29

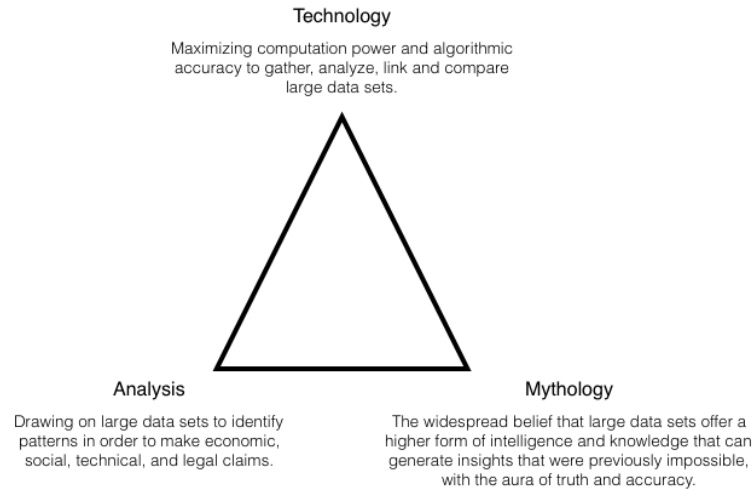with the computational turn in research and thought (Burkholder 1992).

Technology

Maximizing computation power and algorithmic accuracy to gather, analyze, link and compare large data sets.

Analysis

Drawing on large data sets to identify patterns in order to make economic, social, technical, and legal claims.

Mythology

The widespread belief that large data sets offer a higher form of intelligence and knowledge that can generate insights that were previously impossible, with the aura of truth and accuracy.

Fig. 7: *Sleigh, Joanna, Diagram of Boyd & Crawford's conception of Big Data. 2016.*

"We define Big Data as a cultural, technological, and scholarly phenomenon that rests on the interplay of: Technology, Analysis, and Mythology" - (Boyd & Crawford 2012: 663).

Data has an interesting place in our interconnected world. Specifically, big data has been framed within the context of global complexity (Appadurai 1996), the networked society (Castells 2011), information and mobility flow (Urry 2003), and the social formation of data analysis (Ruppert et al, 2013). Rhetoric surrounding this socio-technical reveals that there is an increasing flux of internal cultural debates, as opposed to cultures (Parkins 1978).

Arjun Appadurai's (1996) conception of 'globalization' and non-spaces is particularly interesting for his five 'scapes' are helpful in understanding the constant exchange that is happening of information and ideas across digital and physical forms. As Appadurai notes (1996: 3), data and new technological mediums "*transform the field of mass mediation because they offer new resources and new disciplines for the construction of imagined selves and imagined worlds.*" Appadurai's emphasis on the non-material form and socially constructed nature of space is interesting for it lends itself perfectly to conceptualizing the fluid space that data documents and exists in.

These frameworks of understanding informed the thematic analysis of my research. However overall I took a more emic approach. My involvement in the La Cura program enabled me to use active participant observations as a primary method. This involved direct observation, collective discussions, and the analysis of group produced documents. All of the latter has been collated in field notes and reflexivity journals.

3 Ethical Issues

The method of participant observation, of first hand exploration of research settings with naturalistic field methods, requires vigilant awareness regarding the deception and the absence of informed consent from the people being studied (Dingwall 1980). For this reason, I have remained conscious of the personal responsibility in dealing with issues of awareness. Specifically, from the outset all participants were made aware of my research intentions. Furthermore, in acknowledging the sensitivity of the nature of documenting perspectives, experiences and opinions, the personal information of participants has been masked.

The other ethical pillar of consideration is in regard to the ownership of visuals. Given that many of the mind maps and visuals contained in this research paper were of communal authorship, I have been mindful to correctly reference their authors.

4 Field Notebook

On site at Villa Strozzi, in Florence, my participant observation was structured in accordance to the La Cura program.[4] Maintaining a notebook throughout the week was hence fundamental to my research, as it was an active repository of notes, observations, sketches and texts for future analysis. This was particularly helpful for the communal discussion of theories relating to data along with discussions regarding how we would work with Human Ecosystems. Throughout the rest of the week, this note-taking was enriched with photographic documentation, unstructured interviews, and a communal work diary. Revisiting thoughts and impressions extended these notes and documentations into anthropological process. This journal was openly shown to informants, reinforcing their awareness of the process of documentation.

---

[4] For the purpose of this study, I joined the group that dealt with Data Visualizations and Data Analysis. These individuals dealt with capturing, analyzing and representing the data. For an explanation of the other groups, please refer to Figure 8.

Fig. 8. Baotaz Commons, La Cura Program Structure. 2016, Screenshot.
http://baotaz.net/ (accessed 9 November 2016).

5 Themes / Effects

Through a process of constant comparative analysis, out of the many effects documented as a result of working with big data and data visualizations, two themes emerged as dominant. Firstly, there was a collective agreement that working with big data contributed to a sense of overwhelmingness. Secondly, everyone corroborated the definition that data is a construct inseparable from human choices and perspectives.

5.A Overwhelmingness

The main effect of working with such big data was the sense of being overwhelmed. It was not just the fact that we entered a space of limitless numbers representing an unimaginable amount of things. It was also the large amount of time spent trying to understand the complexities and intricacies of the data, what data was relevant, what data could be accessed, and what forms it took.

"*I thought I knew what data was, I'm not certain any more*" said Christian on day two of La Cura. While another participant in reflection exclaimed "*I want to underline that I was highly frustrated by it… but an interesting and worthwhile struggle.*" The discussions centered on comprehending data and data visualizations and such is documented by the multitudes of mind maps collectively created over the course of the program. Mind map after mind map was made in an attempt to comprehend what data was, what data we could access using the tools, and how we could combine or focus the aggregations. While participants valued this process, at the end of the third day there was a general sense of mental exhaustion. "*The discussions helped a lot to broaden and confuse even more my ideas*" said participant Amy.

Adding to this sense of overwhelmingness was the comprehension

of how much time and what level of skills were required for proper analysis and visualization. "*Not having a proper background in computer science makes you see one hundredth of the spectrum*" for it "*takes way much more effort than I thought… it requires skills and intelligence way beyond average.*" (Interview with Christian & Max). The majority of participants and myself had no background or training with data, and so once the tools and processes of analysis and representation were understood it became more and more of a pressing issue that we did not have the time to achieve what we wanted given the short period of the program. For despite the ease of use of Human Ecosystems as a program, to mine, clean, compile and present data in an interactive visual that can be comprehended, takes more than a few days.

Finally, we must also take into account the situational factors that contributed to this sense of being overwhelmed. The climate was hot, with 35 - 37C days. This was combined with an intense program schedule where attendance stretched from 9am until 8pm. Language was also a factor. As the program was 'international' English was designated as the language of discussion. However, English was not the mother tongue of the majority of participants. This resulted in a lot of miscommunications, slow discussions, and frustrations with communication.
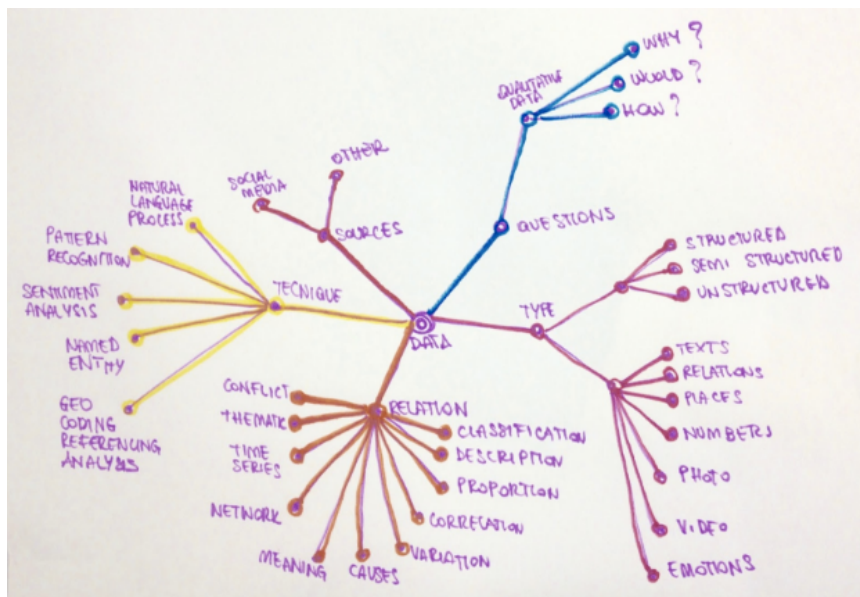

Fig. 9. *Sleigh Joanna, Documentation of conceptualisation of 'Data' mind map. 2016..*
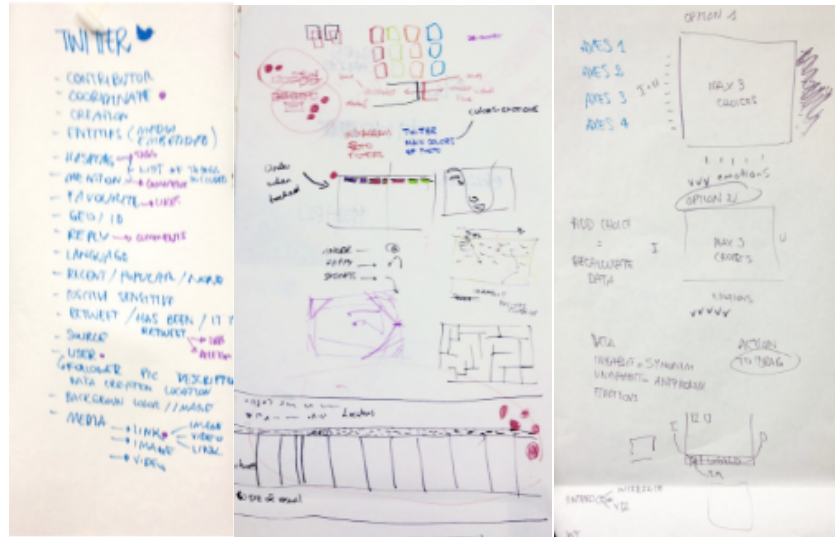
Fig. 10. *Sleigh Joanna, Documentation of social network anaylsis & information visualization sketches. 2016.*
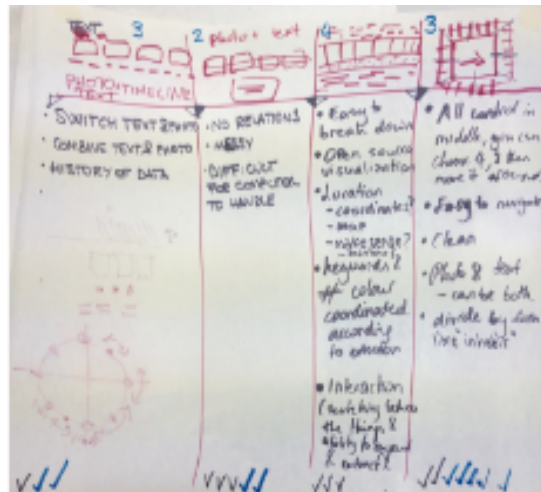

Fig. 11. *Sleigh Joanna, Documentation of voting process for which visualization should be made. 2016.*

5.B Data is an opinion

> "Data is an opinion, seems like a joke, but nevertheless it is a good starting point." La Cura Participant.

From the very first day, a central topic of discussion, one that triggered talk and thought, was the perspective that data is an opinion. "I*t was mind consuming but also deeply interesting to see how one thing, that I generally consider a given, could lead to endless discussion*" one participant reflected. To clarify, this concept referred to the fact that big data can be perceived as a quantification of the world. As a quantification, it is a construction of the human mind and so is not necessarily objective or accurate. Gitelman (2011: 3) observed, data need not be imagined as data in the first place, and this process of the imagination of data entails an interpretative base.

The experiences of the La Cura participants who spent time using the data mining & analysis tool Human Ecosystems reinforced this theoretical questioning. It was their role to make decisions about what

social networks data was collected from, what attributes and variables from these social networks would be counted, and which will be ignored. A process which was inherently subjective. Daniel, one of the researchers explained "*how we define and gather and analyze or otherwise process « data » defines a certain reality in a given context.*"

The big data we collected that week required time, skills and comprehension which showed us all that data, at every point of its existence is a social formation. "*It's the kind of work that gets better with work.*" For even though "*data gives us a view of our reality,*" data requires humans to "*enable people to understand what they see…*" (La Cura Participant Interview, 2016). This is highlighted in 'Statistical Practice: Putting Society on Display' which states that statistical work "*is social through and through*" (Mair, Greiffenhagen & Sharrok 2015: 21).

The understanding that data was an opinion further brought up questions of truth and authenticity. "If data is an opinion, how can it represent things truthfully?" "What about private social media accounts, what is the data not able to show?" "Is the emotional analysis really effective?" "How truthful is the data?" These were just a few amongst a cascade of questions brought up over the course of five days.

Fig. 12. *Baotaz Commons, Data Visualisation Presentation Slide. 2016, Screenshot.*
http://baotaz.net/ (accessed 9 November 2016).

6 Visual Representation

"Man is an animal suspended in webs of significance he himself has spun." - Webber (Geertz 1973: 5).

To communicate the effect of working with data an interactive data visualization was necessary. As one of the main understandings from this research was that data requires human's to make it comprehensible, to give it a perspective, I felt it was important that the visual be interactive, thereby highlighting that data requires human interaction.

Using an open code sourced from Open Processing[5] I created an interactive visual made of the key phrases drawn from the semi-structured interviews. These sentences were co-selected with participants and were chosen as the most succinct references to dominant themes. The first thing the viewer sees is a paragraph of text. With the participants we chose the sentences which.

The moment the user moves their mouse the text deconstructs, creating a chaotic picture of lines and letters. The end result of this chaos is an alphabet categorization of the text. Reflecting the deconstructive and reconstructive processes in data analysis. A further set of manipulations are possible by pressing the alphabet and numbers on a keyboard, making every individual experience different. This represents the subjective nature of data interpretation.

Fig. 13. *Sleigh, Joanna, Visual Representation - text. 2016.*

Fig. 14. *Sleigh, Joanna, Visual Representation – text ordered into letters. 2016.*

Fig. 15. *Sleigh, Joanna, Visual Representation – Variations in chaos. 2016.*

Fig. 16. *Sleigh, Joanna, Visual Representation – Variations in chaos. 2016.*

Fig. 17. *Sleigh, Joanna, Visual Representation – Variations in chaos. 2016.*

6 Conclusion

> "Effective democratization can always be measured by this essential criterion: the participation in and access to the archive, its constitution, and its interpretation" (Derrida, 1996: 4)

Data is part of our life. From social media interactions, health records, phone logs, government censuses and genetic sequences, massive quantities of information are constantly being produced by and about people, things and their interactions. Data is founded on the digital traces that we humans are producing. These traces are also indicative of something else. As the   Lac Cura program demonstrated an underlying essence of Big Data is 'interconnectedness'.

Manovich (2011: 10) identifies three categories of interaction with Big Data: "*those who make it (both consciously and by leaving digital footprints), those who have the means to collect it, and those who have expertise to analyze it.*" These later two groups are a lot smaller, and often the more privileged (Crawford and Boyd: 2012). These inequalities must be examined and questioned for they are structures which produce bias in data and types of research. Tools like Human Ecosystems and programs like La Cura are therefore highly important, for these undermine the divided access, engaging all phases of Big Data, from production and collection to analysis.

Huge power and complexity is inherent in big data and analysis (Boyd & Crawford 2016) and the fact that data may leave us with a sense of being overwhelmed, as the study showed, highlights an awareness of this. Also demonstrated is the huge gap in the capabilities of the average person to access and comprehend the extensive diversity of data processes, types, and forms of representation.

Suchman (2011) via Levi Strauss observes 'we are tools'. Therefore, we must consider how the tools as we use them, participate in shaping the world. We must continue to ask questions. Where does the data come from? Who it is for? is it ethical? How representational is it really? For the era of data has only just begun and it is imperative that we question the assumptions, values, and biases of this new wave research.

References

Csiksznntmihalyi, M. (1990) "*The Body in Flow*" In *Flow: The Psychology of Optimal Experience."* 94 - 100, Harper Perennial.

Appadurai, Arjun. (1996) *Modernity at large: cultural dimensions of globalization*. Minneapolis, MN: University of Minnesota Press.

Baotaz. "website with La Cura documentation 2016". Last Accessed November 2016. http://baotaz.net/abaot-az/.

Boyd, Danah, and Kate Crawford. (2012). Critical Questions For Big Data. *Information, Communication & Society.* 15, no. 5: 662-79. doi:10.1080/1369118x.2012.678878.

Burkholder, L. (1992). *Philosophy and the Computer*, Boulder, San Francisco, and Oxford:Westview Press.

Castells, Manuel. (1996). *The Rise of the Network Society*. Malden, MA: Blackwell Publishers.

Derrida, J. (1996). *Archive Fever: A Freudian Impression*, trans. Eric Prenowitz, University of Chicago Press, Chicago.

Dingwall, R. (1980). "Ethics and ethnography", The Sociological Review, 28, 871-891.

Gitelman, L. 2011. Notes for the Upcoming Collection 'Raw Data' is an Oxymoron, [Online] Available at: https://files.nyu.edu/lg91/public/ (accessed 23 July 2011).

La Cura Summer School. "La Cura Summer School - EN." Last accessed November 2016. http://la-cura.it/summerschool/eng/.

Kranzberg, Melvin. (1986). "Technology and History: Kranzberg's Laws", *Technology and Culture* 27, no. 3 (1986): 544-60. doi:10.2307/3105385.

Mair M, Greiffenhagen C and Sharrock WW. (2013). "Social studies of social science: A working bibliography", National Centre for Research Methods Working Paper Series, No. 3219.

Manovich, L. (2011). Trending: "The Promises and the Challenges of Big Social Data", Debates in the Digital Humanities, ed M.K.Gold. The University of Minnesota Press, Minneapolis, MN <http://www.manovich.net/DOCS/Manovich_trending_paper.pdf>.[accessed 15 July 2011].

Posner, J., Russell, J. A., & Peterson, B. S. (2005). "The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology", Development and Psychopathology, 17(3), 715–734. http://doi.org/10.1017/S0954579405050340

Ruppert E, Law J, Savage M. (2013). "Reassembling social science methods: The challenge of digital devices". Theory, Culture & Society. 30(4): 22–46.

Suchman, L. (2011). "Consuming anthropology", in Interdisciplinarity: Reconfigurations of the Social and Natural Sciences, eds A. Barry & G. Born, Routledge, London, [Online] Available at: http://www.lancs.ac.uk/fass/doc_library/sociology/Suchman_consuming_anthroploogy.pdf

Unwin, T. (2009). *Information and Communication Technologies for Development*. Cambridge: Cambridge University Press.

Urry, John. (2003). *Global Complexity*. Malden, MA: Polity.

Geertz, Clifford. (1973). *The interpretation of cultures*, New York, Basic Books.

Interviews

Amy W. "La Cura Participant." in discussion with the author. August 26, 2016.
Christian J. "La Cura Participant." in discussion with the author. August 26, 2016.
Jeremy. "La Cura Participant." in discussion with the author. August 26, 2016.
Kirsten. "La Cura Participant." in discussion with the author. August 26, 2016.
Max K. "La Cura Participant." in discussion with the author. August 26, 2016.
Peter B. "La Cura Participant." in discussion with the author. August 26, 2016.
Sharon L. "La Cura Participant." in discussion with the author. August 26, 2016.