

Datatopia – Data Tracking in an Alternative world

Yumeng Wang
Parsons School of Design, BFA
Design and Technology
New York, USA
wangy594@newschool.edu

ABSTRACT

Datatopia is an interactive installation that reframes the unsustainable online data tracking system into a creative sci-fi narrative. It intends to inform the audience the vulnerability of their data, collected by marketers for immoral commercial profits. Datatopia consists of an interface for experiencing the narrative of traveling in a world made of text and logs every bit of information the audience has input to generate a unique printout story that they can take away with them. Datatopia examines the relationships between us and the data we generate on daily basis, and speculates the future definition of personal privacy.

Author Keywords

Data tracking; privacy; ad blockers; surveillance; natural language processing

ACM Classification Keywords

Design; experimentation

US, OUR DATA

How many stories can be generated from just looking at one's data without actually knowing the person? Even the objective meta data of a person are able to create subjective narratives. The linkability of different types of personal data: bank information, online shopping preferences, emails, messages, etc., ties the knots between scattered pieces of information.¹ All the information together is not as simple as aggregation, but implies more aspects of us than we are aware of. We are the ones that produce our data, but the otherness in these data can also distort parts of us. Instead of absolute ownership, other kinds of relationship between us and our data should be developed in order to understand the consequences of the otherness in these data and how they might betray us.

Data visualizations reflect on the patterns of certain characteristics of data effectively. For personal data, visualizations have been used commonly to record one's life events. Japanese artist On Kawara has an abundant collection of works dedicated to analog personal data keeping. In his piece *I Went*, On Kawara draws the paths of places he has went to everyday on a map. While his other piece *19,834 Days* marks every single day of his life as yellow and blue dots. Graphic designer Nicholas Felton has created a series of

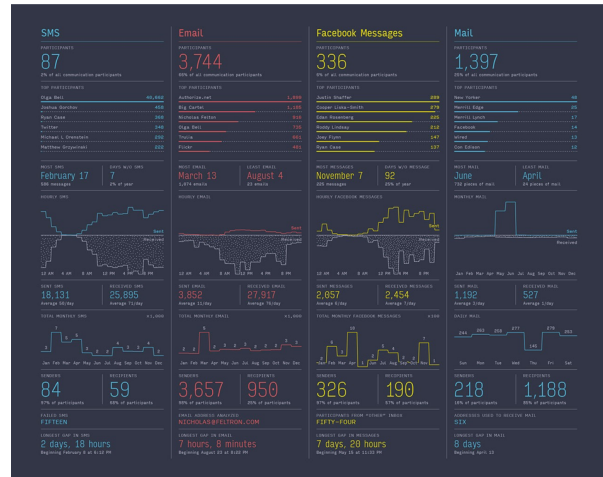


Figure 1. Nicholas Felton, Personal Annual Reports

personal data record - Personal Annual Reports. Almost eerily detailed, Felton's graphical reports keeps tracks of the numbers of his emails, SMS, Facebook messages, mails and analytical information from these data. In this process, people almost turn themselves into machines that record their every activity. To some extent, the intentions of the activities have shifted. How much of them are conducted naturally or for the sake of being recorded?

With the extensive data tracking system embedded in our digital presence, our data can tell different stories from our intention. Commercial products based on our online data are targeted back to us like ghosts wherever we go in the browser. Our decisions are constantly obstructed by their influences, and constructing personal identities has become almost impossible without the interference.

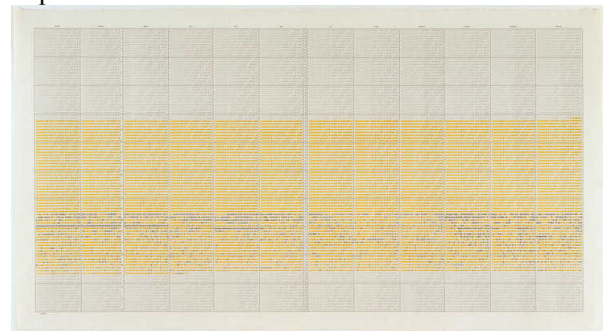


Figure 2. On Kawara, 19,834 Days



Figure 3. Ghostery



Figure 4. Blendle

Ad blockers and several alternative online publications have made the effort to minimize the effectiveness of online advertisements. Different from popular ad blockers such as AdBlock and Privacy Badger, Ghostery² drills through the fundamental problem of ad technology: the publishers do not obtain profits directly from the content they create, so putting ads along with their content becomes a forced act. Instead of selling user data to any data mining marketers, Ghostery sells them merely to the publishers themselves, ensuring their benefit as well as protecting users' privacy. New journalism platform Blendle³ removes all ads but charges a small amount of fee for reading each article voluntarily. It believes that if the reader really enjoys an article, he or she should be willing to make a minor contribution for the content creators. However, if the article turns out not worth reading or is accidentally clicked, the user still has the option of having their money back.

Not only in ad technology, but also in computer vision and many AI's, people are examined as data objects carrying patterns that indicate intentions. These processes are loops that do not involve humans. The red and green squares often seen in computer vision images are shown to help humans understand the computer's thinking, while the algorithms running through the images make them operational rather than representational. Humans are the targets in these operations, but ironically decisions are largely relied on algorithms than subjectivity, creating the delusional truth that machines understand the emotional human beings. For instance, an AI might confuse a wedding reception and a terrorist attack because they have

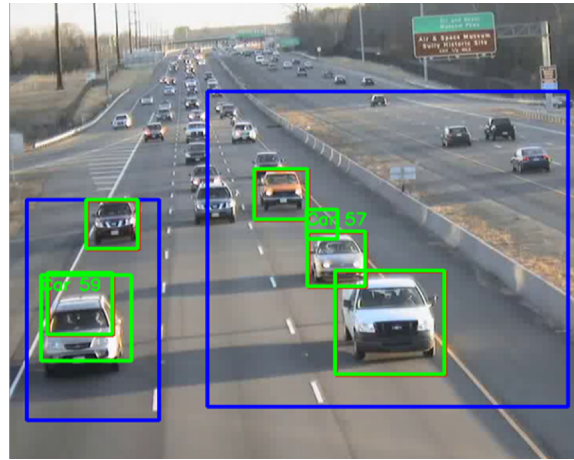


Figure 5. Computer Vision

similar patterns of a large amount of people gathering. A large amount of victims from false drone attacks have suffered from the overconfidence of algorithms. The fear of subjectivity has shaped our data into potential enemies against us.

A BRIEF HISTORY OF PRIVACY

The overflow of personal data in these relationships also becomes harmful to personal privacy, whose definition has shifted radically over different stages of civilizations. Privacy is considered unnatural in many theories, because in ancient societies such as Greece and Rome, most activities are conducted in public space and privacy almost does not exist. People tend to choose wealth and convenience over having private activities. Privacy has developed slowly for the past 3000 years until the recent two centuries, moving beyond physical boundaries to psychological awareness. The rise of individualism and more possible recording techniques of our data have led to the recognition of privacy.

In 1844, when Italian politician and activist Giuseppe Mazzini was exiled in London, he believed that the British post office had been opening his mails for information of an insurrection in Italy. The Mazzini affair was considered "the first modern attack on official secrecy".⁴ In this case, Mazzini reckons that his mails are his privacy, while the government is convinced that the mails are secrecy, the opposite of the transparency that our society has always asked for. The paradox of our expectations for both transparency and privacy will inevitably lead to some forms of compromise.

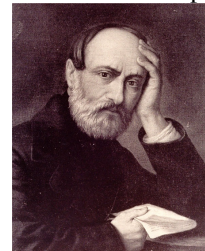


Figure 6. Giuseppe Mazzini

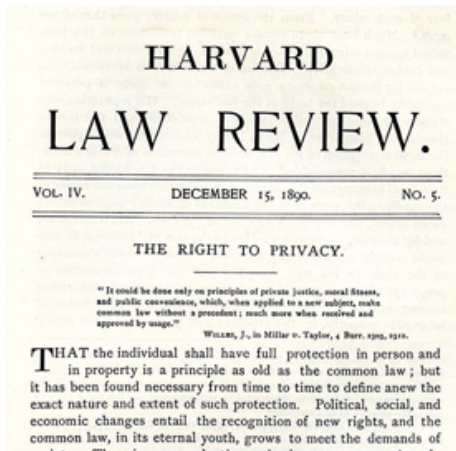


Figure 7. The Right to Privacy, Samuel Warren and Louis Brandies

In 1890, two Boston lawyers Samuel Warren and Louis Brandeis published “The Right to Privacy” in Harvard Law Review.⁵ It was considered the first declaration of the right on privacy in the US, even though it didn’t obtain much attention at the time. In the statement they condemned the media’s disclosure of personal information. “The press is overstepping in every direction of the obvious bounds of propriety and of decency. Gossip is no longer the resource of the idle and of the vicious, but has become a trade, which is pursued with industry as well as effrontery.” They also pointed out that what should be public instead are the actions from government departments.⁶ Privacy became a right to fight for.

More recently, Snowden’s disclosure of PRISM is an agreement that allowed major technology companies to share their users’ information with the government. Last year, CISA, or Cybersecurity Information Sharing Act, has passed the Senate in October and signed into law in December. CISA is a federal law aiming to “improve cybersecurity in the United States through enhanced sharing of information about cybersecurity threats, and or other purposes”. These legal regulations have made personal privacy more vulnerable than ever.⁷

Other obvious instances related to privacy protection that we encounter frames privacy into commodities and even luxury. Major online publications Forbes and Wired have kept users using ad blockers from viewing their websites. Alternatively, they can pay for the service to remove ads from their viewing content. The domain hosting company GoDaddy has similar restrictions when you sign up for their services. Paying extra fee allows your contact information to be protected. In 2015, AT&T charges 29 dollars more for their gigabit fiber-to-the-home service, with the additional protection of web browsing from being tacked.⁸

In the future, the meaning of privacy will change again as the surveillance structure evolves to be more intricate. Governmental surveillance and corporate

Here’s The Thing With Ad Blockers

We get it: Ads aren’t what you’re here for. But ads help us keep the lights on. So, add us to your ad blocker’s whitelist or pay \$1 per week for an ad-free version of WIRED. Either way, you are supporting our journalism. We’d really appreciate it.

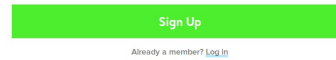


Figure 8. Wired Ad Blocking Page

surveillance will bond even closer. Our voluntary participation will seamlessly provide abundant personal data in this three-dimensional surveillance system.⁹

INTENTIONS

Investigating the implications of various relationships between us and our data, as well as where the future of personal privacy leads, Datatopia replicates the online data tracking system in an alternative world. It still tracks the user’s information but only for benign purposes: generating a sci-fi story. The instructions and questions are carefully curated to provide clues that all the choices are made have consequences in the end. The printouts of the stories draw more attention from the audience to read closer to what their data have become. The stories are created with techniques of natural language processing and generative text. These techniques embody the comparison of algorithms versus subjectivity discussed in the relationships between us and our data.

GENERATIVE TEXT

Generative text emerged from the analog era when Dadaists cut out individual words from a piece of text and randomly arranged them to form new texts. This cut-up technique was popularized later by poets and writers including William Burroughs and Brion Gysin, in their work *The Third Mind* in 1977. Found in 1960s, Oulipo (*Ouvroir de littérature potentielle*) also implemented methods to reduce subjectivity in writing and constrain writing techniques with mathematical references.¹⁰ For instance, one of their most popular rules was called N+7. For any types of text, replace the nouns with the seventh noun in the dictionary following the original word. Another rule was Snowball, in which each word has to be longer than the previous word. The participants of Oulipo were not only writers, but also mathematicians.

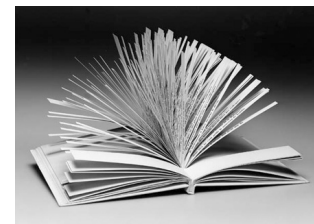


Figure 9. An Example of Oulipo

```

epoch 1.01 / 50 - 10 / 17889 - loss = 3.09551
epoch 1.01 / 50 - 11 / 17889 - loss = 3.09921
epoch 1.01 / 50 - 12 / 17889 - loss = 3.10262
epoch 1.04 / 50 - 13 / 17889 - loss = 3.09947
epoch 1.04 / 50 - 14 / 17889 - loss = 3.10271
epoch 1.04 / 50 - 15 / 17889 - loss = 3.10455
epoch 1.04 / 50 - 16 / 17889 - loss = 3.09896
epoch 1.05 / 50 - 17 / 17889 - loss = 3.10133
epoch 1.05 / 50 - 18 / 17889 - loss = 3.10517
epoch 1.06 / 50 - 19 / 17889 - loss = 3.09817
epoch 1.06 / 50 - 20 / 17889 - loss = 3.10124
epoch 1.06 / 50 - 21 / 17889 - loss = 3.10079
epoch 1.07 / 50 - 22 / 17889 - loss = 3.10241
epoch 1.07 / 50 - 23 / 17889 - loss = 3.10439
epoch 1.07 / 50 - 24 / 17889 - loss = 3.10485
epoch 1.08 / 50 - 25 / 17889 - loss = 3.10351
epoch 1.08 / 50 - 26 / 17889 - loss = 3.10317
epoch 1.08 / 50 - 27 / 17889 - loss = 3.10674
epoch 1.08 / 50 - 28 / 17889 - loss = 3.10921
epoch 1.08 / 50 - 29 / 17889 - loss = 3.10754
epoch 1.09 / 50 - 30 / 17889 - loss = 3.10634
epoch 1.09 / 50 - 31 / 17889 - loss = 3.10984
epoch 1.10 / 50 - 32 / 17889 - loss = 3.10445
epoch 1.10 / 50 - 33 / 17889 - loss = 3.10887
epoch 1.10 / 50 - 34 / 17889 - loss = 3.10521
epoch 1.11 / 50 - 35 / 17889 - loss = 3.10181
epoch 1.11 / 50 - 36 / 17889 - loss = 3.10755
epoch 1.11 / 50 - 37 / 17889 - loss = 3.10749
epoch 1.12 / 50 - 38 / 17889 - loss = 3.10156
epoch 1.12 / 50 - 39 / 17889 - loss = 3.10393
epoch 1.12 / 50 - 40 / 17889 - loss = 3.10527
epoch 1.12 / 50 - 41 / 17889 - loss = 3.10952
epoch 1.12 / 50 - 42 / 17889 - loss = 3.10762
epoch 1.13 / 50 - 43 / 17889 - loss = 3.10664
epoch 1.13 / 50 - 44 / 17889 - loss = 3.10770
epoch 1.13 / 50 - 45 / 17889 - loss = 3.10612
epoch 1.14 / 50 - 46 / 17889 - loss = 3.10670
epoch 1.14 / 50 - 47 / 17889 - loss = 3.10181
epoch 1.14 / 50 - 48 / 17889 - loss = 3.10311
epoch 1.15 / 50 - 49 / 17889 - loss = 3.10165
epoch 1.15 / 50 - 50 / 17889 - loss = 3.10634

```

Figure 10. Training a Recurrent Neural Network Model with Torch-rnn

Algorithms for generative text have evolved in the digital era. A few lines of code in Javascript can easily create a function that analyzes the concordances of every word in a piece of text. Simple techniques with statistics methods such as frequency-inverse document frequency (tf-idf), Markov Chain, and context-free grammar, are also interesting enough to generate a large amount of text without context, which embodies the principles in early analog generative literature. To add context and even mimic the subjectivity in natural language, machine learning methods such as recurrent neural networks (RNN) and recursive neural networks, have brought algorithmically generated text much closer to simulating human’s writing techniques. In these methods, models are trained with a huge amount of data sets to learn how words and phrases are placed in relation to each other. This process is time consuming even for high stand equipments, and even for open-sourced RNN training such as Torch-rnn, the Lua script takes about an hour to train a Shakespeare excerpts of a few thousands words with the GPU on a personal computer.¹¹

The original intention of analog generative text is to avoid human interference, however, it goes to the opposite direction when computational algorithms try to simulate human’s natural language. These algorithms can easily generate a 10,000-word novel, but these texts are not for us to read; they are better just to think about.¹²

USER JOURNEY

Resembling an old-schooled TV, the installation is a turquoise and yellow box with a screen in the front, a slot below it for paper to come out from the printer inside, and a mouse and a keyboard. The start screen shows the caption Datatopia and asks for a username to begin the experience.

Miyeon walks to the installation and decides to spend some time looking at it closely. After inputting her username, Miyeon is informed that she has signed up for an interstellar trip to Datatopia, a planet on the edge of the Galaxy of Emotions. Two other passengers Rory and Yue, who have travelled there before, will accompany her during the trip as the tour guides. Although Miyeon can choose some small characteristics for the other two travelers, such as hair color, their names are actually the previous two



Figure 11. Datatopia Logo



Figure 12. Datatopia Setup

users’ names. Through this detail, it implies that other people have used this project before and more information from them will unfold as the journey continues.

Before the spaceship launches and during several scenes later in the experience, Miyeon will be promoted to answer questions related to ad blockers, social media and online browsing habits in order to precede the narrative. The spaceship travels through the Galaxy of Emotions, stopped by a few other planets to pick up “emotions”, and eventually lands on Datatopia at nighttime. Every object on Datatopia is made of words and sustained by new words from the tourists. It will not exist if no one comes to visit.

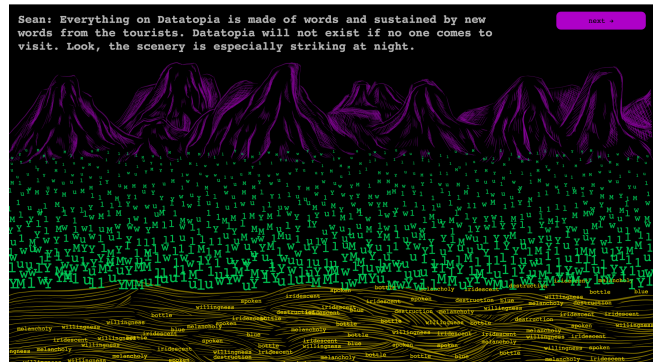


Figure 13. Datatopia, Landing Scene

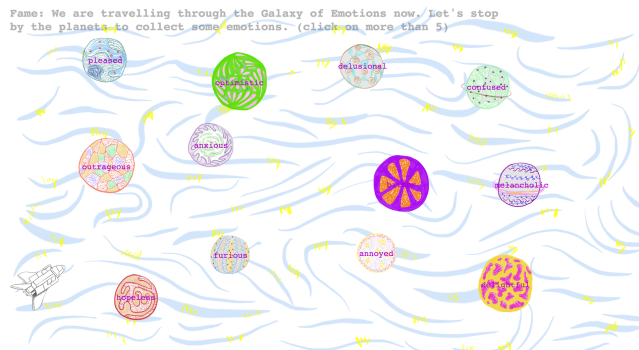


Figure 14. Datatopia, The Galaxy of Emotions

Even the air on Datatopia is made of words. In this scene, Miyeon can control a giant lung with the cursor on the screen to breathe “air words” into the lung. After these initial reactions on the new environment, Miyeon arrives at a dark road paved by short questions and with street lamps on two sides of the road. Answering each question will light up the corresponding lamp until all lamps are lit up and together they light up the day. At the end of the road, there is a deep hole with mysterious force that drags the rocks to form a special pattern. As other objects, all the rocks are also made of words. When Miyeon throws rocks into the hole to see the pattern, they also choose the words, with or without intentions. The pattern created by the rocks is a number as the passcode to open a box with a question mark on the top. It contains every tourist’s traveling record. When it is opened, Miyeon is able to view her own unique record, constructed by the experience she has just gone through with details written based on their choices and answers in each scene. When she hits the button “view your record”, a physical copy of the traveling experience is printed out from the bottom of the installation.

USER TESTING AND REFLECTIONS

As an installation piece in an exhibition space, Datatopia takes about 7 minutes or longer to go through the entire experience. Although the instructions in each scene are reduced to minimal length, it is still a text-based narrative that requires users’ patience. In user testing without my presence at the piece, around 40 people have interacted with Datatopia in the given three hours, and among them, 23 people have finished their stories and received the printouts. The interaction flow goes more smoothly than expected, especially with people who have more interests and even discuss with their peers before entering their answers to the questions. Those people also tend to read their printouts carefully, even though 80% of the story is algorithmically generated and not intended for contextual meanings. Most reactions of the printouts are when people see their names along with two other previous users’ names on the second line, they feel rewarded about this customized and unique story generated from the effort they just put into the digital experience. But surprisingly, 3 people interpret the almost random texts as fortunetellers, reading them out loud to their peers.



Figure 14. Datatopia User Testing

Although the overall feedback is positive in terms of the interaction and graphics, the weak points are the connections between each two scenes, especially in the later part of the storytelling. The scene with the deep hole and rocks gives the hint to the one-digit passcode in the next scene, where the user is asked to put the passcode. However, about a third of the audience doesn’t enter the correct number the first time, because the connection is not strong enough to tie in these two scenes. Adding transitions between them or highlighting this part of the instructions might be an effective solution.

CONCLUSION

Datatopia does not necessarily answer the questions of developing healthier relationships with our data or how privacy will evolve in the future, but it expects the convoluted data tracking system to reduce the unethical branches. Marketers should not make general assumptions that when users browse the Internet, they are always buying things. Genuine and meaningful content deserves our attention and financial contribution. But this can only happen in an environment without the pollution of bad advertisements and interfaces that trick users to make irresponsible decisions.

REFERENCES

1. Citizenfour, Laura Poitras, 2014.
2. Ghostery, <https://www.ghostery.com/>
3. Blendle, <https://launch.blendle.com/>
4. Lepore, Jill. “The Prism”. *The New Yorker*. 2013. <http://www.newyorker.com/magazine/2013/06/24/the-prism>
5. Warren, Samuel, and Louis Brandeis. *The Right to Privacy*. Harvard Law Review, Vol. 4, No. 5, pp. 193-220, 1890.
6. Dorothy J. Glancy, "The Invention of the Right to Privacy", *Arizona Law Review*, v.21, n.1, pp. 1-39, 1979.

7. Greenberg, Andy. "Congress Slips CISA Into a Budget Bill That's Sure to Pass". Wired. 2015.
<https://www.wired.com/2015/12/congress-slips-cisa-into-omnibus-bill-thats-sure-to-pass/>
8. Brodtkin, Jon. "AT&T charges \$29 more for gigabit fiber that doesn't watch your Web browsing". Ars Technica, 2015.
<http://arstechnica.com/business/2015/02/att-charges-29-more-for-gigabit-fiber-that-doesnt-watch-your-web-browsing/>
9. Lyon, David. *Theorizing Surveillance: The Panopticon and Beyond*. Cullompton: Willan, 2006. Print.
10. "A brief guide to Oulipo". Poets.org. 2004.
<https://www.poets.org/poetsorg/text/brief-guide-oulipo>
11. Torch-rnn, jcjohnson. Github.
<https://github.com/jcjohnson/torch-rnn>
12. Coburn, Tyler. "My Life in the Cloud". Motherboard. 2015. <http://motherboard.vice.com/read/my-life-in-the-cloud>