Dietmar Offenhuber: Collaboration Via the Many Traces Our Data Sets Leave Behind

Article *in* IEEE Computer Graphics and Applications · August 2020

DOI: 10.1109/MCG.2020.3007058

3 authors:

Dietmar Offenhuber
Northeastern University
91 PUBLICATIONS **1,081** CITATIONS
SEE PROFILE

Bruce Donald Campbell
Rhode Island School of Design
45 PUBLICATIONS **644** CITATIONS
SEE PROFILE

Francesca Samsel
University of Texas at Austin
84 PUBLICATIONS **619** CITATIONS
SEE PROFILE

Uncorrected Preprint Jul 1 2020. Published version at https://doi.org/10.1109/MCG.2020.3007058

Dietmar Offenhuber: Collaboration via the Many Traces Our Data Sets Leave Behind

Deitmar Offenhuber
Northeastern University

Bruce Campbell
Rhode Island School of Design

Francesca Samsel
University of Texas

Abstract – We asked Dietmar Offenhuber for an interview upon investigating his recent works, Staubmarke *(Dust mark) and* Ozone Tattoo*, which together suggest an artist with an insightful eye for injecting particulate data engagingly into art pieces that spur awareness in an audience. We were impressed by his deep considerations on the potential for scientific and informatics data sets to offer potential for, and foster, wider discussions and collaborations.*

Francesca: Your background spans a wide range of disciplines from the visual arts and architecture to science and technology. As I explored your work I was struck by the seamless integration of the disciplines from your background and their manifestation in the work. Can you talk about your overall development process, given the multifaceted roots that underpin projects like *Ozone Tattoo*?

Dietmar: In terms of my recent interest, I develop projects that let me investigate the material underpinnings of data. What counts as data is a trivial question from a computational perspective, since data always come in the form of digital records. The question becomes less trivial when we look at how data are generated. For example, archaeologists may look at a buried artifact as a datum; geologists may look at geological strata as natural archives. In my work, I am interested in such material traces and their relationship to digital information. In my view, the material origin of data is somewhat of a blind spot in data visualization. Visualization begins with a data set that already exists; it is mostly silent about how these data were captured from the world. The visualization, therefore, cannot capture all the facets of the research.

In *Ozone Tattoo* (Figures 1 and 2) and *Staubmarke* (Figure 3), I approach this issue from the perspective of pollution sensing. We usually discuss particulate matter and ozone pollution through numbers that were presumably captured by well-calibrated sensors, but pollution also leaves traces everywhere in the environment. Fine dust accumulates on all surfaces of the city, ground-level ozone is, unlike stratospheric ozone, harmful for all living beings and leads to visible damage to plants. I am trying to make these traces visible and legible, to provide context and sensory grounding for digital sensor values. In the case of *Staubmarke*, I washed visual markers into the patina of the city to allow observing the accumulation of dust.[1] This technique is known as “reverse graffiti” — painting by cleaning off the dirt. You can only examine the dust once you see it in contrast with a clean surface. Over time, the markers will

[1] See https://offenhuber.net/project/staubmarke-dustmark/

fade away, conveying a sense of how quickly the dust accumulates.

In the case of *Ozone Tattoo*, I exposed the leaves of sensitive plants to ozone to create damage patterns that serve as a visual reference.[2] The project expands on the practice of "ozone gardens" which include ozone indicator plants such as tobacco, green beans, and milkweed. In both projects, the markers function similar to the legend of a map, they allow comparison and focused observation of a phenomenon in the world.

F: *Ozone Tattoo* uses the surrounding atmosphere to record the data and create the visualization itself.

D: I describe this as *autographic visualization* (a term I came up with together with my colleague Orkan Telhan).[3] In the 19th century, the idea that nature inscribes itself was very popular. Everything that happens in the world leaves traces somewhere. Today, this idea can be found in the forensic imaginary; that we can reconstruct the past from innocuous traces. In the context of pollution, such traces and their evidentiary value have a social and political dimension. For example, communities may struggle to show that they are affected by pollution, even when official data show that everything is fine. Data artists often work with pollution data — they plot it in different ways, showing inherent patterns and their aesthetic qualities. They don't often think much about how these data were recorded. I am also interested in aesthetics, but on a more fundamental level — how pollution or global warming manifest themselves in the world on a sensory level. I am not criticizing aesthetic explorations of digital data sets, they can reveal a lot of meaningful and surprising patterns, but I pursue a different angle, seeing autographic visualization as somewhat of a counter model of data visualization.

---

2 See https://offenhuber.net/project/ozone-tattoo/

3 See Offenhuber, Dietmar. 2019. "Data by Proxy — Material Traces as Autographic Visualizations." *IEEE Transactions on Visualization and Computer Graphics*. https://doi.org/10/gf649f.

Uncorrected Preprint Jul 1 2020. Published version at https://doi.org/10.1109/MCG.2020.3007058

Figure 1. *Ozone Tattoo* image with caption

F: Let's revisit your thoughts about exploring data that shine a light on information collected beyond the original intent.

D: When you work with physical traces, there are so many layers that are not visible. A digital data set, according to Claude Shannon, contains a finite amount of information. A physical trace is much richer and contains almost unlimited amounts of it. *Ozone Tattoo* is connected to a whole research field into bio-indicators and proxy data sources. Of course, I cannot show all of that. I don't want to make autographic visualizations that are so complicated that you have to read a manual. The notion of sensory evidence is important – that you notice a pattern and learn how to interpret it. The second goal is to emphasize the material nature of data — that all sensors are essentially autographic instruments, and their data are physical traces that are in some way still part of the phenomenon.

F: The intriguing thing is that the work has so many ties back to what the data is about – its physical aspects are what makes it so stunning. There is the context for data, but you go one step further towards the discussion of how visualization affects people instead of just how it effects them.

Figure 2. *Ozone Tattoo* image with caption

D: I became interested in the notion of data as a physical trace during the *Trash Track* project, which was my doctoral dissertation at the MIT senseable city lab.[4] We attached location trackers on individual pieces of garbage in order to see where it ends up. The trackers used a mixture of GPS and cell-tower triangulation to acquire the location. We would receive really strange results sometimes such as a heap of garbage ending up in the backyard of a mansion on an island close to Seattle. We said "this cannot possibly be the place", and found out that it was a location artifact that resulted from the localization method of the sensor. The cell-tower on this island was the only source of information for the sensor, which was likely buried on a trash barge passing it. In other words, you cannot understand the data without understanding the materiality of the apparatus and the assumptions baked into it.

In my book "Waste is Information," I explored this idea that garbage is saturated with information, but that information is latent or hidden. And waste does have an affective quality: people are fascinated by the waste system — everyone has opinions and beliefs about how it works and what is wrong with it.[5]

F: Let's talk more about "Waste is Information". Does it build on the *Trash Track* project?

---

4 See http://senseable.mit.edu/trashtrack/

5 Offenhuber, Dietmar. 2017. *Waste Is Information: Infrastructure Legibility and Governance*. Cambridge: MIT Press.

D: The book is basically about how cities, civic activists, and all the other actors involved in urban governance make infrastructures legible through data. At the time, there was much discussion of the promise of big data for urbanism — the idea that we have all this data and can use it to improve cities. By focusing on waste, I explored situations where this premise fails — where data are rare, sketchy, and controversial. I wanted to show the complexity and labor of collecting data about the waste system — an infrastructure that is usually opaque. It's a bit about methodology on how to investigate waste systems from the bottom up, taking a forensic perspective to figure out what is going on in the system. It was published in the infrastructure studies series at MIT press, so it takes on more of a design research rather than an artistic perspective.

Figure 3. Staubmarke *(Dust mark)* image with caption

D: When we work with material information, we have to take a relational view. There is no definition, no legend that tells us what a particular trace means — everything is implicitly expressed in the causal relationships within a phenomenon. I love Richard Alley's book on ice-core research.[6] He is a paleo-climatologist who extracts ice cores in Greenland and reconstructs past temperatures. He makes it clear that there are so many different phenomena even within this single ice column that can be used to reconstruct past temperatures. But none of them give you an absolute point of reference. One has to correlate

6 Alley, Richard B. 2014. *The Two-Mile Time Machine: Ice Cores, Abrupt Climate Change, and Our Future - Updated Edition*. Princeton University Press.

all kinds of material properties both locally and globally: from lake sediments to tree rings, from coral layers to the isotopic ratios of the ocean water. Such a relational perspective of comparing and correlating many phenomena is based on this perspective that everything around us is saturated with information, and people devise different strategies to reveal it.

From the perspective of autographic visualization, I compare Alley's scientific approach with much simpler practices of revealing traces by community scientists and activists. For example, the bicycle advocates in New York City, who document vehicle tracks in snow after a snowstorm to reveal the areas of an intersection that are not needed for vehicular traffic and could be used for bike lanes or parklets instead.[7] What connects these vastly different practices is the idea that data collection can be a sensory and aesthetic process — not a bureaucratic exercise.

F: Data visualization often benefits from contextual information. Are you are saying that there is contextual information existing within the data itself?

D: Many scholars in Science and Technology Studies (STS) argue against the idea that data are universal and context-free and suggests that we always need to consider the context from which it came. Yanni Loukissas explicitly addresses this in his book "All Data Are Local."[8] We should not forget about all these material conditions that shaped the process of data capture. The material context offers important additional information that is often necessary to make sense of a data set.

We can illustrate this point with the popular dataset containing origin-destination data for hundreds of millions of taxi trips in NYC. If we plot the geographic locations of the taxi pickups and drop-offs, we get, for example, a very dense map of the Manhattan street grid. But we might also notice that the map is blurry in certain parts of Midtown and Battery Park. This is a result of high-rise buildings diminishing GPS reception and increasing the error. What is essentially 2D information contains information about the 3D shape of the city. But we can use this information only if we understand the material conditions of data capture. Every data set contains unintentional traces of its own generation: little inaccuracies, default values, transcoding artifacts, etc. Normally, you would exclude all data points that seem wrong. But if you focus on these errors, you can find all kinds of interesting things.

F: You obviously wear different hats to pursue your interest in urban studies. How do you identify which errors to examine further?

D: I work on topics that seem to have little to do with each other, but that are held together by certain themes that I keep exploring. I often come back to older projects where a certain issue raised my curiosity, but I was not able to explore it at the time – because the research had a different purpose, or because it would have made the whole project too complicated. I am a believer in mixed methods – I find it interesting to use a qualitative lens for what are often considered quantitative research problems, and vice versa. Data-driven projects usually

---

7 See https://twitter.com/search?q=sneckdown

8 Loukissas, Yanni Alexander. 2019. *All Data Are Local: Thinking Critically in a Data-Driven Society*. Cambridge, Massachusetts: The MIT Press.

benefit from additional contextualization. Of course, you cannot do everything in a single project. So, I may decide to move on and come back to process from a different angle later.

F: Do you have any current projects that stemmed from examining the errors?

D: A few years ago, I was commissioned to create a spherical animation for the National Museum of Emerging Science and Innovation in Tokyo, Japan (Figure 4). The brief was to bring together global data sets that show the future of megacities cities across the world and their role in the global context. I had difficulties finding the right data sets that cover the whole globe for the phenomena I was interested in. And when looking into the geo-grids that exist, I always came across a single data source that is baked into them—from economic activity to resource consumption. This data source is comprised of global satellite mosaics of the night-lights going back to the 1970s. The history of this "black marble" data set is fascinating because its emergence was mostly accidental. In the 1950s, the US military needed a weather satellite for surveillance flights. But the discovery that these satellites were able to capture the night lights of cities was not anticipated. And then it turned out that these night lights are the best available global indicator for economic activity in a particular location, becoming somewhat of a workhorse for economists. Although this anecdote did not play a role in the animation for the museum, it became an important starting point for further work in data proxies.

These examples, from trash tracking to taxis and city lights, have brought me away from the popular metaphor finding the needle in the haystack, to the insight that all data sets reflect many different phenomena at the same time, intentional or not. Critiques from the humanities often attack the idea of objectivity by pointing to all the arbitrary decisions and assumptions that go into a data set. They show, in other words, that data are like language. In my examples, I have less difficulties with the word "objective" when I translate it as "object-like." And focusing on object-like qualities leads to a different critique, attacking the notion that information, data, ideas etc. are abstract and instead emphasizing the implications of their different embodiments. Of course, data are political, based on worldviews and assumptions, and these assumptions are embodied and enacted through objects. If you will, I take an ontological perspective on data and the question, "what is a data set?"

Figure 4: “Sorting out cities” exhibit at the Museum of Emerging Science and Innovation in Tokyo (Miraikan). Image courtesy Miraikan

F: Can you elaborate on what constitutes a data set?

D: I am currently teaching a course at Princeton on environmental sensing in its broader material, social, and political implications. We are looking at how evidence is constructed in environmental justice conflicts — especially in the current context of the EPA under the Trump presidency. We are looking at how agencies monitor and enforce violations, how community scientists produce different kinds of evidence and use it for advocacy.

In one exercise I describe as “Sensor Autopsies,” students disassemble a sensor and investigate how it works, what it responds to, and what it cannot represent. This conveys a better sense of what a data set can and cannot show. In another exercise, students search for proxy data sources that can be used to measure a pollution phenomenon without digital sensors. This provides a sense of how data relate to the larger phenomenon. Both exercises are closely related to the practice of community scientists such as Public Lab[9] or the Louisiana Bucket Brigade,[10] who document and visualize environmental burdens together with affected communities by building sensors and collecting sensory evidence.

Environmental laws typically specify threshold values that decide whether pollution is problematic or not. This means that the circumstances of sensing, the locations of stations are a political issue. But even when the levels of multiple pollutants are below their legal limit, taken together, they can still create a big burden. The tendency to fragment and isolate variables for air, water, soil pollution does not represent how an issue is experienced. To

9 See https://publiclab.org

10 See http://www.labucketbrigade.org

address this issue, community scientists combine sensor data with other forms of documentation, including subjective ones such as odor logs. In all big environmental disasters, it came to light that people had complained for years about smells in the air or the drinking water. So for me, these recorded sensory experiences are an important form of data.

F: I am curious about how you instill in your students the practice of looking 'around the corner' at the data.

D: I sometimes avoid providing too much information up front. I want students to be able to discover the information for themselves. In one course, we started looking at EPA enforcement data, which has countless legal terms attached. Students struggle with that, but this is an essential part of understanding environmental data, especially in combination with other perspectives, including the sensor autopsy and the exploration of proxy data sources.

B: Your perspective reinforces my belief that we don't think enough about the collaborative value of data and how if others knew how data was going to be collected, they might point out a small change to the collection process that would be of enormous value to future needs from which a collaboration could then thrive.

D: I think the collective aspect is crucial because evidence construction is also a social process that involves showing, exchanging, debating, and persuading. Community science is a good example of how data capture can be a form of collective experience, from figuring out ways to document an environmental issue and produce evidence to the collective sensemaking of what has been collected. My colleague Sara Wylie at Northeastern works on a method to capture emissions from fracking with photopaper.[11] The physical traces of the stained paper become a way for the affected community to experience the causality behind an invisible issue. It becomes much easier to contextualize, and if necessary, challenge environmental data collected by public agencies.

This form of phenomenological sensemaking is becoming more popular also outside the environmental domain. I was recently surprised to hear a group of artificial intelligence researchers describe their approach of mapping deep learning networks in very similar terms. They treat the fully-trained neural network like an unknown natural phenomenon, map it in "activation atlases,"[12] and probe its behavior—almost like 18th Century natural philosophers categorizing clouds.

B: As artificial intelligence continues to grow in public awareness, I can anticipate your perspective and work on data traces may provide a useful context for discussion. Do you think you'd do work to foster that consideration?

D: I am interested in novel research approaches that look at data from a different angle. In the current public discourse of the COVID pandemic, I found the role of autographic visualizations

---

11 See https://publiclab.org/wiki/hydrogen-sulfide-photopaper

12 Carter, Shan, Zan Armstrong, Ludwig Schubert, Ian Johnson, and Chris Olah. 2019. "Activation Atlas." *Distill* 4 (3): e15. https://doi.org/10.23915/distill.00015.

Uncorrected Preprint Jul 1 2020. Published version at https://doi.org/10.1109/MCG.2020.3007058

in addressing questions such as “should I wear a mask?” or “how should I wash my hands?” very interesting. We have seen videos where soap has been replaced with black ink to show the effectiveness of different hand washing techniques or an experiment where the potential contamination in a buffet is visualized with invisible ink fluorescing under blacklight. Lab technicians have coughed into Petri dishes with and without masks, leading to an impressive colony of bacteria in the latter case. Also popular were Schlieren imaging videos that show the distance exhaled air travels. While the traditional data visualizations of epidemiological data focused on the discovery and interpretation of patterns, the autographic examples focused on building evidence, often directly addressed at “conspiracy theorists.” I think that it is no longer sufficient to publish a data set and expect the public to accept it as uncontroversial facts. And these recent examples show that researchers start thinking about how to communicate their methods, not just their results.