Data Is Political: Investigation, Emotion and the Accountability of Institutional Critique

Bio

Amber Frid-Jimenez and Ben Dalton are co-directors of *Data Is Political: On Contemporary Art, Design and the Politics of Information*, a research project that brings together artists, designers and data scientists in a discourse about the aesthetic and political dimensions of visualizing immense archives of digital information. Amber Frid-Jimenez is an artist, an associate professor at Emily Carr University of Art & Design, and a principal of AFJD Studio, an interdisciplinary design firm at the intersection of architecture, information and ecology. Ben Dalton is a principal lecturer at Leeds Metropolitan University, a PhD candidate at the Royal College of Art researching new forms of digital public space in the CX Lab, and a visiting professor at the Bergen National Academy of Art & Design. They began their collective work eight years ago at the MIT Media Lab.

Introduction

Analysts speculate about the impact of big data on international business and government policies. Meanwhile, proportionally few designers and artists have yet to take up the challenging questions of how large-scale digital information systems will reshape our future. A small but growing community of technically savvy designers and activists are leading the charge, raising aesthetic and political questions on the perception, use and sometimes misuse of data – questions traditionally left to science, journalism, and politics. Despite the efforts of a growing few, the need for more designers and artists to keep pace with the rate of technological innovation remains no less urgent. By bringing artists and designers together with data scientists and policymakers within the larger project of Data Is Political we catalyze debate about the role of design and art to produce meaning through the presentation and analysis of big data.[1] Our goal for this chapter is to lay the foundation and vocabulary on which that discussion builds and accountability is maintained.

We use the term 'big data' to refer to enormous scale data storage, processing and connectedness, often on global scales. Digital information systems have reduced the cost of copying and connecting cultural, scientific and social signals. In contrast to the high production costs of analog copies, once digitised, all kinds of archives from early movies to rooftop radiation sensor streams to political voting records are reduced to signals that can be duplicated and transferred fluidly across our global internetwork at a relatively low cost and without the slightest

[1] Data Is Political is a research project directed by the authors that investigates the aesthetics and politics of data, their collection, visualization, and distribution. The project explores the recent context of ‘big data’ and its cultural implications for artists and designers. Big data is a term used to describe the expanding archives of digital information collected using the internet, mobile phones, surveillance cameras and other sensors in the environment. The project addresses the aesthetic and political dimensions of the contemporary condition that arises from an increased focus on collecting and presenting data. The project takes the form of interviews, symposia, projects, articles and a forthcoming book publication.

degradation. Designers and artists use these digital signals to remap the original data to new forms in service of creating new meanings. In the contemporary data-saturated context, the process of remapping signals from colour to sound, from percentages to spatial plots, from individual images to curated stories becomes as effortless as drag-and-dropping bits or writing code commands.

Key technological breakthroughs in computation over the last century precede the current ease with which we produce, present and distribute data. By 1948 Claude Shannon's information theory had proven mathematically that digitised files could remain perfect copies through repeated processing and global transmission (Shannon 1948). Our current modes of production take advantage of the rapid rate of duplication and sharability that Shannon's theory affords. Moore's Law, which describes the doubling of transistors on circuits every two years, has characterized the acceleration of storage and networking technology, reflecting the ongoing trend of rapid information infrastructure growth from the invention of the integrated circuit in 1958 onwards.

Despite these key innovations, until recently working with data was an elite activity. The technology required to capture and re-edit film, or to collect and map weather data was out of reach of all but a few well-funded scientists, commercial entertainers, advertisers and military researchers. Information processing tasks were expensive and required specialized training. A politics of scarcity limited the need for large-scale accountability. We define 'accountability' in this context to mean an obligation to demonstrate clearly the methods by which data are collected, processed, used, presented and distributed. Until data processing moved off institutional computers and into the public around 1999, identifying who was collecting and using which data and how was relatively easy because the community of researchers who had access was small enough to monitor themselves.[2] However, as the tools of digital copying continue to reduce in cost, more and more people and organisations have access to larger and larger datasets, calling for an evaluation of modes and practices of accountability.

The scientific community has long worked with data representation, establishing widely accepted conventions of accountability, including open data, reproducibility, peer review and explicit statements of uncertainty. Recently, designers have been called upon to work on interfaces that represent or are driven by huge data sets. However, in large part, designers have not been trained in scientific or statistical data conventions, and so often have not considered how their design choices, such as which filters, visual shapes and simplifications to use, affect meaning and application. The reach of big data and big processing is making a much wider group of people, including graphic designers and artists, into data scientists without them necessarily

[2] Distributed.net, launched in 1997, was the first system to enable a coordinated distributed data-analysis task, but it wasn't until 1999 that the distributed method was popularized. The Space Sciences Laboratory at the University of California, Berkeley launched SETI@home (Search for Extra-Terrestrial Intelligence at home) allowing users to participate in a distributed computing task that analyzed radio signals to search for signs of extra-terrestrial activity and intelligence.

realising it. The creative potential of this new situation is boundless, and necessitates the development of new conventions of accountability for data-driven design practices.

Traditionally, data has been used to hold people accountable for their actions, identifying hidden connections, corroborating stories, providing evidence and proving theories. In this sense, data remapping and visualization has always been a technique of accountability. But, data are also used for many other reasons now – to connect us, evince emotion and mediate our experience of each other and the world. Artists and designers engaging with data provide new possibilities for storytelling and communication, but must also keep in mind that as this artistic field grows, data resources grow even faster, and so too the power of the people that own them. Creative practitioners should focus attention equally on designing new accountability tools recognizing that companies with big data centres and infrastructure continue to manage and restrict access, even as we design more tools for data accountability. Our responsibility then is to question the limits of current accountability tools, providing new techniques and interfaces to ask how (and where) data resources are managed.

In this chapter we describe data visualisation and other forms of art and design working with data as a process of remapping. We define three axes of data remapping: *investigative design*, *emotive reframing* and *institutional critique,* on which we lay out the implications of big data on design practice. Reviewing the full potential of tools for accountability requires us to imagine them being used on our behalf as well as against us. We analyse the role of big data in accountability by examining both its potential benefits and its uses as a means of *unaccountability*, or evading practices of accountability. We conclude by exploring strategies to design out unaccountability, arguing for the potential of principles of openness to guarantee access and use of big data accountability tools.

Axes of Remapping

> It's important to realize that visualization is a medium. You are framing your argument all the way from how you gather your data, to how you curate it, to what comes out … you are framing reality. – Fernanda Viegas[3]

Our understanding of the potential uses of big data within accountability technologies is grounded in a history of exemplars. However, the significant examples of data remapping were developed at a time when only comparatively small data sets were technically possible. The potential of big data tools is still largely untested, and represents a significant space for exploration by artists, designers and data scientists. Many data scientists and designers we spoke to within the last year as a part of the Data Is Political project testified to the need for testing, discourse and convention within this relatively new territory. Philip DeCamp, data

[3] All of the quotes in this chapter are taken from conversations recorded by the authors with artists, designers and data scientists and from presentations at the Data is Political symposium in Bergen. These perspectives were collected between 2011 and 2013. http://dataispolitical.net

scientist from the Cognitive Machines Group at the MIT Media Lab described his process:

> As you start dealing with more complex visualizations shoving together multi-modal sets and adding a bunch of camera movements and scripted events, you have to make a lot of design choices. If you're looking at new kinds of data, there is no convention; you just have to make things up.

On one hand, the artistic freedom DeCamp describes is tantalizing for applied artists who dream of fewer constraints on their work, but, on the other, indicates the need for a deep examination of how certain design decisions affect meaning.

The term 'data visualisation' often describes contemporary tools and techniques of storytelling with data. We find it unsatisfying in its implied limitation to the visual, proposing instead a definition that includes many overlapping forms from different fields and perspectives. Our conversations have yielded descriptions of data visualisation ranging from a single, personalised sentence summarising political news, to spatial and auditory systems, to tools for interrogating the visual effect of filtering world health datasets as a means to provide evidence to policymakers. From this breadth, we define the three axes of remapping outlined below.

Investigative Design

Investigative design is grounded in a history of visual language for statistical analysis and scientific investigation. Although not constrained solely to visual media, the definition of 'graphical excellence' from Edward Tufte (1983) characterises the objectives of investigative design well, as complex data communicated with "clarity, precision and efficiency" . Tufte's intention is to enable an audience to ask question of complex systems with a minimum of distortion and distraction in the design elements. As the speed increases with which data can be collected and processed, we expect greater levels of investigative design within journalism. The New York Times for example have dedicated 'graphics editors' who work with data and visualisation in their journalism.

> The goal is still to see how I can help guide my reader, how I can explain something … The reporter writes a sentence, and then the supporting evidence is the quote below it. We write a sentence and the supporting evidence is the chart below it. – Amanda Cox

Design that provokes questions enhances accountability within complex systems, making clear ties in networks that previously have been hard or impossible to see.

Emotive Reframing

Whereas investigative designs often have specific tasks of understanding as a core motivation, emotive reframing images tend to be focused more on persuasion. Emotive reframing in design resonates with someone through poetic analogy or experience. Although not always the case,

reducing the functional efficiency of investigative design often can heighten emotive reframing. A distinct authorial voice defines the aesthetic and narrative aspects of this type of practice. Media theorist and activist, Florian Schneider uses the analogy of framing, an act traditionally used in photography and conceptual art practices, to describe how this mode of data-driven design functions:

> This moment of framing tells us more about the desire to establish a relationship rather than about reality itself.

The degree of reframing can vary from close correlation to abstract representation. A single emotive image can be used to represent an entire data set. For instance, Casey Reas described the example of abstracting out the analysis of climate science data on gradual warming to an image of a polar bear on a melting ice flow. In this example, the photograph acts as a hook, communicating the issue through the immediacy of a single image without the need to spend time interpreting analytical details. However, to remain part of the data visualisation process, emotive reframing must link back to more detailed investigative design or discussion.

Combining emotive and investigative dimensions in a single design is a challenging balance of storytelling and interrogation. Philippe Rekacewicz, radical cartographer and journalist for Le Monde Diplomatique, combines emotional impact with the precision of traditional map making. Many of Philippe's maps are hand drawn, an approach that captures the human stories of the underlying data, while at the same time acknowledging his own authorial voice within the constraints of geographic mapping. For example, *Les Trois Frontières de L'Europe* maps immigration camps and the European frontiers to people seeking asylum (Rekacewicz 2006). The map efficiently shows the infrastructure of migration and in that sense conforms successfully to cartographic convention. But, Rekacewicz's map goes beyond convention, using the uniqueness of a hand crafted line to emphasize the humanity of the migration: the ambiguous borders of internment camps outside the Schengen Area are sketched in wider, less certain pencil line, and dominating black circles and blood-red type mark the brutality of the lives lost in perilous sea crossings. Design that reframes data to resonate emotionally with an audience has the potential to reach beyond already engaged groups to a wider population, bringing to light injustices and mobilise efforts to enforce accountability.

Institutional Critique

The axes of investigative design and emotive reframing capture most examples of visualisation. However, a third form of remapping data exists. This third axis measures the meta-analysis of the systems of data collection, processing and presentation themselves. While traditional techniques, such as source citation and error bars that express uncertainty, capture part of the overall frames of reference, assumptions, and ideological underpinnings of data science, self-criticality is lacking from many examples of investigative design.

Artists have been interrogating the use of materials as well as critiquing institutional

organisations for a long time. We can therefore draw key examples of institutional critique in data analysis from art history. *MoMA Poll* by Hans Haacke is a classic example of institutional critique (1970). For the piece, Haacke placed a transparent voting box in the gallery to measure gallery audience opinion on MoMA donor and board member Governor Nelson Rockefeller's support for President Nixon's Indochina Policy in 1970. Due to the political controversy that the piece elicited, rumor has it that the day after the opening the work came under the threat of removal. The artwork commented directly on the relationship between the gallery and global political systems, implicating Haacke himself, his curator and the institution in a web of political and social power structures. The threat of removal endangered the artist livelihood, and constituted a clear and real form of self critique and, by extension, institutional critique. Questioning assumptions and perspectives within a system surfaces fallacies and prevents people and organizations from being able to intentionally hide their motivations.

> Archives are far from neutral repositories of information. They have become vibrant sites for cultural production. The archive is no longer seen simply as a static form to be mined, but instead is vital for staging new agendas in politics. – Nomeda & Gediminas Urbonas

Unaccountability Technologies

The greatest challenge to accountability technology is not technical or aesthetic, but instead concerns the politics of accountability. There are many motivations to remain unaccountable. The dimensions of remapping described in the previous section each offer ways to be accountable and at the same time can be used to deflect or repress accountability. Greater storage, processing and connectedness motivate some people to use big data to create systems to block accountability, while calling others to action for the sake of holding those with power over the infrastructure accountable for their decisions and actions.

Designers are often employed to use their insights and skills in persuasion to the advantage of political and commercial ends. Investigative designers leverage the visual language established by science and journalism documentation because of its power for explanation and exploration. Often, established indicators of clarity and efficiency can be appropriated to lend a sense of authority or trust to misleading data. For example, the formality of graphs, maps and other typical visualisation techniques are often assumed to be inherently accountable. However, information used to create a design may have been fabricated, the underlying process may have been modified, or a distorting assumption may not have been made clear. As data size and processing complexity increase, audiences run the risk of becoming dependent on shorthand symbols of trust in place of taking the requisite time to investigate the sources and assumptions themselves.

> Because we usually deal with 'data' and 'numbers' people think that visualisation is a 'neutral' tool: "Well, I'm just showing the numbers, it's the truth". – Fernanda Viegas

Of the three axes of remapping that we have outlined, emotive reframing is the most dependent

on the author's choices to ensure the connection between the underlying information and the story being told. This means that a skilled storyteller can easily use the techniques of emotional connection and persuasion to deliberately draw attention away from certain questions or ideas and toward others. With the advent of big data tools of profiling and real-time processing, personalisation can be used for unaccountability to adjust stories being told for each person individually, to match their prejudices, or to take advantage of their particular blind spots. In the remaining part of this chapter, we review the potential of design to support accountability through big data remapping, to resist or identify unaccountability, and to sustain tools of accountability in the face of complexity and opposition from those who do not stand to benefit from them.

Designing Out Unaccountability

Successfully balancing the axes of remapping creates design that is investigative, emotive and critical. Citation of sources, expression of uncertainty and critique of process are key features of effective data-driven design. The audience should be assisted in asking what range of possible meanings the source data suggest, about the fallibility of the narrator, and about the assumptions and simplifications that have been made. These general principles offer a framework to tackle unaccountability in big data systems and underpin guidelines for artists and designers who find themselves dealing with questions of data remapping without the context of traditional statistics or visualisation training. Data science too can grow more effective from this dialogue by borrowing institutional critique from the arts, and extending criticality beyond the current data analysis to include the broader system of funding, research direction and infrastructure resource. Furthermore, emotive reframing can be employed to reach wider audiences with greater impact.

Distinguishing between visualisation norms that infer symbolic trust and the accountability of an underlying data process requires a level of information design literacy.

> Giving people the literacy to understand is a key concern to us. It is part of the empowerment. If people are literate, and they understand how these things work then that gives them the power. – Martin Wattenberg

Data literacy must attempt to keep pace with adapting forms of decoy and camouflage used to evoke a sense of trust without supporting analysis. When designed effectively and with an eye toward literacy, big data systems enable 'training' examples to be added to many more facets of life. Experience at reading information design will develop as the tools to process and display data become more prevalent.

Digital display along with associated connectedness and processing, allow for infinite layers of detail in a visualisation. In print design, finding space to show both an overview and detailed story on a single page was a considerable challenge. While the 'no wasted ink' design principles are still a core aspect of investigative design, dimensions such as zoom, linked space, transparency, blur and time allow for complexity to be available to a viewer if needed. Interaction

goes a long way to displaying the effects of the assumptions and filters that have been applied.

> With the kind of data sets we are working with now having a single result is very misleading … you need to be able to compare multiple results, that's really where the information is … in the ability to filter the data in different ways to reveal other possibilities. – Casey Reas

The viewer can play with the assumptions of a visualisation through interactivity in order to understand their effects. However, not everyone will have time or inclination to explore underlying process in data remapping. Systems should record the reviews of those that do, so that people can rely on personal networks of trust.

Beyond systems that expose key parameters to interactive control are toolkits that enable entire processes of remapping to be explored. Growing ecosystems of computation, rendering libraries and tools are enabling designers, programmers and artists to build new forms of remapping. A trade-off must be struck between complexity of the interface and depth of possible exploration. While such systems require significant training and experience to use, the number of people with access to this potential is much greater than in the time of data and processing scarcity. The number of data remapping designers will continue to grow as the community of experienced users continue to document their development and use of these tools. Free and open access therefore must be a requirement for software ecosystems of exploration, documentation and sharing of data remapping techniques.

A Race for Openness

Open data, open tools, open processes; without these, how can we audit the stories that are being told? Data collection, processing and remapping technologies should be, at a minimum, open source so that the assumptions used can be analysed. For incremental exploration and development of these tools, a free software license is also required so that creators with significant resources do not retain control over what can and cannot be held accountable with those tools. Freedom to appropriate also decentralises the remapping process by removing a need to seek access or permission.

> The question of how [big data systems are] controlled becomes essentially the question of politics. The good news is that we've been working on this problem for several tens of thousands of years and the downside is … politics isn't a solved problem, and so the politics of data is not going to be a solved problem either. – Benjamin Mako Hill

The scale of big data systems presents new challenges to open access approaches to accountability. If a large corporation or government were to open up its archive for download, no one but other corporations and governments would have the resources to store and process it.

> Everything goes in the direction of opening data … If there is no redistribution of

> infrastructure, it will sound great, but who is actually able to access that data?
> – Marcell Mars

Instead, in the current ecosystem of data infrastructure, individual users and small external organisations have to rely on the big data provider to maintain an 'open API' (Application Programming Interface) for them to query parts of the data when they need them. Or people might use data analysis tools hosted by the big data provider to do their processing and remapping. The difficulty is that if tools rely on cloud services, they become dependent on or sustained by commercial interests. Controlled proprietary systems and limited APIs can shape what questions can be easily asked, or stories can be told. Using closed APIs limit how visualisations can be notated or augmented with alternative analysis.

Closed centralized information systems can lead to unaccountability and corruption when the actions and behavior of the few people who control them are far from public view. Peter Sunde, co-founder of the Pirate Bay, points out the potential dangers of giving control to a concentrated few:

> All sorts of power corrupts people, so we should consider that when we build structures, that if you centralise things, corruption is possible.

Decentralization, on the other hand, avoids concentrations of power that may lead to corruption. Indeed, open decentralized systems that put power in the hands of a multitude can engender accountability by their nature. However, decentralized systems are not free from other problems. Decentralizing information can obfuscate the critical power structures underlying systems of governance. For example, recent neoliberal literature such as David Beito's *The Voluntary City* argues in favor of decentralizing areas of government control such as the oversight of the U.S. financial markets and environmental projection (Beito, Gordon, and Tabarrok 2002). The ability to act freely avoiding accountability in this case is retained through a sleight of hand where a centralized objective is coordinated to maintain unaccountable control of a seemingly distributed system. That is to say that distributing bits of information here and there in an uncoordinated way produces a signal to noise ratio that makes it difficult to analyze the whole system. A centralized information system, on the other hand, is perhaps easier to scrutinize from below as long as it is accessible. While in this view centralized information systems seem to provide the potential for simple and effective accountability, adequate accessibility may be too difficult to guarantee. Decentralized systems can attempt to avoid centralized control by encouraging a diversity of organisational architectures and contributor perspectives, and fostering institutional critique.

Big data infrastructures could potentially enable unconstrained critique and accountability of the big data corporations and government departments themselves. Decentralised systems offer one possible approach to this problem. The enormous library of popular culture stored and made accessible in bittorrent ecosystems practically demonstrates the ability for big data storage distributed across the home computers of a large audience of users.

> What bittorrent did for the internet was fundamentally amazing. It made computers servers again. It made computers converse with other computers. – Max Van Kleek

Project like WebBox demonstrate the potential for distributed big data to offer “a new type of web where people can communicate with each other personally” (Van Kleek). Decentralised systems would be self sustaining once started because much of the data held by big data corporations is currently collected from the work of the audience. For example, in a social network site the users add most of the value in analysing and documenting their social connections. Decentralised systems where an individual’s data are stored in a distributed way across a network of other people’s computers has been and still is technically possible. However, the weakness of decentralised systems of accountability is that any system can be 'gamed', and the limits enforced to control this manipulation slow the effectiveness of the operation as a whole. Intentional system sabotage from coherent centralised power is a significant threat. Moreover, the reason this kind of system does not compete with large-scale corporate cloud services is economic and political.

Those people and organizations who want maintain unaccountability and who have vested interests in profiting from the work of the users within closed systems will resist a move towards open big data systems for accountability. In the face of such resistance, we need to identify, collate and critique attempts to derail progress towards further democratisation of data analysis and open systems of accountability. Making open systems that use big data to ensure accountability requires sustaining sufficient accountability in closed systems. This race condition produces an urgency to build open systems before the scale of data collected and processing available in closed systems is too great to compete with.

Building structures to document aspects of control in closed systems and openly cataloguing unaccountability technologies is imperative. This documentation must present dimensions of network connections, influence, ownership, physical resources and bottlenecks in the system. The three axes of remapping (investigative design, emotive reframing and institutional critique) offer a framework to enable artists, designers and data scientists to tackle the challenge of designing and building the open tools and systems necessary to maintaining accountability and shaping how big data will affect the way we communicate, share resources and live in the future.

References

Beito, David T., Peter Gordon, and Alexander Tabarrok. 2002. The Voluntary City: Choice, Community, and Civil Society. University of Michigan Press.

Haacke, Hans. 1970. MOMA Poll. Installation.

Rekacewicz, Philippe. 2006. “L’Europe Et Ses Frontières Paradoxales - Les Blogs Du Diplo.” November 27. http://blog.mondediplo.net/2006-11-27-L-Europe-et-ses-frontieres-paradoxales.

Shannon, Claude E. 1948. “A Mathematical Theory of Communication.” Bell System Technical Journal 27 (October): 379–423 and 623–656.

Tufte, Edward R. 1983. The Visual Display of Quantitative Information. Graphics Press.