

4S 2022 talk

Let's contrive a simpler analysis than the analysis of the perverse queerness of analysis, one clearly toward the far end of our positivist-interpretivist spectrum: "Environmental activist Diane Wilson is badass." Let's look for some data, to assess its rigorousness. Using Tor Browser I searched DuckDuckGo (this is as rigorously distanced from my usual browsing patterns as I could get) for the string of characters "diane wilson barn". This was the first result:

And here's the data that link takes you to:

I am glad to have helped make this data not simply available, but easily *discoverable*. And by "data" I mean this photograph of Wilson in her barn in Port Lavaca, Texas, taken by Emree Weaver for *The Texas Tribune*.



Wilson is a long-time environmental activist and self-described "[unreasonable woman](#)." She *looks* kinda badass here, doesn't she? Still, we should check if more data is available, since the rigor of my interpretation is in question.

The stuff in Diane's barn in the photo is not data; it is just stuff, although there is enough of it to call it Big Stuff. This Big Stuff is Big Data's future anterior, to get all fancy and French about it.<sup>1</sup> You could call it "raw data," if you prefer to get all scientific about

---

<sup>1</sup> Data on the future anterior is available in this databyss: <https://returntocinder.com/motif/futureanterior/sources>

it, but I think that would be an inferior choice, if a popular and indeed hegemonic one. The photograph is data because, unlike the stuff in Diane's barn, [it is available and discoverable, and it is available and discoverable because it has metadata](#). Hence a first proposition, in various expressions, on our way to establishing rigor:

(A1v1) Data if and only if metadata.

(A1v2) If metadata, then data.

(A1v3) Data is anything with metadata.

(A1v4) Only meta/data.<sup>2</sup>

From which it follows logically that:

(B1) Availability if and only if meta/data.

But by the additional proposition –

(C1v1) Meta/data if and only if archive.

–we can then conclude that

(C1v2) Availability if and only if archive

Availability of data *requires* that there be a place prepared for it, a place that *someone* must have readied and now maintains. (This by the Law of Meta/data Hospitality and its Domestic Care Clause, which I have proven elsewhere.)

Given (A1), (B1), and (C1), then by the Law of Archive Fever (also proven elsewhere):

(D1) Availability IFF *archon* (power/authority/ruler/State/G-d/SysAdmin).

analyzed the cultural formations in and through which we differentiate and confer value on our knowledges and knowledge-making practices—ordering them between the poles of positivism and interpretivism, let's say, since we are all in a hurry—alongside the cultural formations through which we differentiate and confer value on sexualities and sex-making practices, between a norm and its perversions. These two orderings, my analysis tried to show, took on similar patterns, produced by the immaterial but real forces of a cultural field or discursive and social structures. That analysis led me to conclude that (I am summarizing) there are neither Two Cultures nor One, as well as

---

<sup>2</sup> I could have written an essay on data and data science entirely around marks like “/” that are (absolutely) ~~fundamental~~ to (post)structuralist language ideologies yet annoy the shit out of anyone operating with a representationalist language ideology (which includes all of us, to some meaningful degree). More importantly but not unrelatedly, data science has difficulty accommodating and working with slashed entities that do not coincide with themselves, concepts under erasure, parenthetical traces of meanings, disseminated and transmuted senses, etc. etc. TL;dr: computers don't (yet) queer very well and “sense,” as renowned logician Charles Dodgson has shown, is essentially queer, or at least curiouser and curiouser.

both One Culture and Two. I signified this paradoxical state with the formulation “ $\sqrt{2}$  Cultures,” to index its position outside an integral system.

In other words: I did not think I was going to be writing about data availability in terms of perversion, but now that I am, I see how it makes sense and hope to convey that to you. What might we say about the availability, analysis, and sharing of (qualitative) data if we framed interpretivism less as signature mode or method of a distinct knowledge domain, and more as perverse positivism? Interpretivism, perhaps, as positivism’s perverse style, form, figuration, performance? More importantly: how can we design digital infrastructure (a word which does not appear in the workshop summary paper) to take advantage of such perverse relations? Can our research infrastructures support not (only) an ethos of reproduction and reproducibility, but an ethos of deviation, experimentation, and surprise?

“The shattering effect of [sexual] perversion,” contends Jonathan Dollimore, “arises from the fact that it is integral to just those things it threatens.” Teresa de Lauretis extends this analysis of the “constitutive paradox” of perversion to characterize it as “both central and yet disruptive; necessary and yet objectionable; a ‘deviation’ from the norm and yet more compatible with positive social goals...regressive or involutory and yet expressive of an original intensity of being.” (De Lauretis 1994, 25) It’s precisely these “perverse dynamics,” these integral shatterings, that I think are at work and in play in *all* systems of data archiving and analysis, interpretivist and/or positivist alike.

#### METHODS AND MATERIALS

A digital research infrastructure (the [Platform for Experimental Collaborative Ethnography](#), or PECE) was designed and built using entirely open source software (Drupal-based) by the investigators (including the author) and their consultants on [NSF grant #1535888](#) (2015-19), to serve as digital research infrastructure and data management system for the research project “Environmental Health Governance in Six Cities.” Bugs revealed themselves and were fixed; versioning occurred. More bugs were revealed and fixed; more versioning occurred. A significant amount ( $p < 0.05$ ) of hair was pulled.

Additional research collaborations were co-constructed (see Supplementary Data, Table 1) over the next few years to provide additional data and cultivate deep expertise on data

sharing practices and theories in the empirical humanities (see *infra*). PECE was *bricoleured* into an experimental system (Rheinberger XXX) to better understand how digital infrastructure *structures*—from an ambiguous location below and/or beyond and or/within (*infra*)—structures of knowledge, rather than simply supporting them.<sup>i</sup> These methods and hypotheses are consistent with results from research on other infrastructures, from databases to petrochemical plants to disease classification systems to sewage treatment (see e.g. Larkin, Bowker and Star).

The Research Data Alliance ([rd-alliance.org](http://rd-alliance.org)) was joined by the investigators in 2013, first as fieldsite for the study of the diverse data sharing practices, cultures, and infrastructures of the ostensibly positivist sciences, eventually as participant members as co-chairs of the ostensibly interpretivist Empirical Humanities Metadata Working Group. Narratives were constructed and delivered, repeatedly, explaining that 1) references by data scientists to “social science” or “qualitative data” were insufficiently granular and otherwise inadequate and unproductive; 2) at least some cultural anthropologists did not self-identify as social scientists, even if they appeared to share certain practices and concepts as well as ethical concerns such as privacy, permissions etc.; but 3) neither should they be categorized or ontologized as “digital humanists,” even if they appeared to share certain other practices and concepts, primarily because of their continual generation of new data through often idiosyncratic processes of selection, collection, interviewing, hasty scrawlings of free [sic] associations in field diaries, etc.

The [National Bridge Inventory makes important data available](#) concerning the state of these infrastructural structures, findable and accessible as interoperable ASCII files. Normal data, and in this case, normal data is good data. My former student (and Lead Platform Architect of our PECE Platform) Lindsay Poirier requires students in her “Intro to Data Studies” class at UC Davis to find that available data, download it, and begin to work with it. They quickly find that that data is only meaningful as data because the Federal Highway Administration also makes available numerous documents, such as [Revision of Coding Guide, Item 113 - Scour Critical Bridges](#), detailing the evaluative judgments beneath, after, within, or simply *about*—some of the possible readings of the *meta-* of metadata-- the coded values in

the data set. Her students must then avail themselves of this necessary supplement, one among numerous other such documents and data sets that together constitute an extensive and elaborate disseminatory structure of metadata, the interpretive apparatus enveloping even the most ostensibly positivist data and data systems. These interpretivist moves are both central and disruptive to positivist structures, a “‘deviation’ from the norm and yet more compatible with positive social goals.”

Now you may say that such interpretations of bridge data is not what we really mean by interpretation. I find such constraints perverse, but whatever; as feminist psychoanalyst Muriel Dimen notes, “Perversion may be defined, after all, as the sex that you like and I don’t.” Real interpretation, some anthropologists will object, is what happens when my self, as an embodied research instrument, encounters people and situations in all their near-ineffable complexity and nuance, residing with them over lengthy periods of time, and then interpreting out of that rich context into a published text likewise shaped by subtle connections to rich literatures and a complex intellectual genealogy. No metadata can adequately capture that and make it, too, available.

To which our reply is: maybe so, but why don’t we try? Why not undertake a few experiments, undoubtedly crude at first but surely refineable and extensible, to see how annotations and related metadata (infra)structures can serve to create this kind of scholarly provenance, archiving it and making it available along with other data?

This is what we are working on and towards. The real perversity is: we can’t really know if our elaborate, expensive, and often frustrating and hair-pulling efforts will in fact be worth it. But this is the perverse risk of the experimental style in general, even in its most positivist forms: if it is in fact experimental, the outcome is uncertain.

Part of such a shift is a related perversion of availability, where the key concerns and questions pertaining to *any and all* data are not matters of degree or quantity, but matters of (infra)structure and style: *what* data is made available, and how is it made available—through what structures and relationalities? In other words: it’s the metadata, stupid.

So I am excited about any initiative that makes qualitative data shareable, like the Qualitative Data Repository, and think it a great thing that these are now multiplying. And we are happy to see an increasing (albeit still relatively small) number of cultural anthropologists become data-curious, work and advocate hard for more of that openness, and have designed PECE to support the goal of making as much new ethnographic data available as possible within ethical limits, situationally constituted. The first step is providing that metadata-structured place where an ethnographer can place the interview she just recorded and/or transcribed, the scan of the piece of ephemera he picked up at last week's clinic, the field sketches they drew in their notebook from the shareholder's meeting.

(Even this kind of availability involves a lot of really hard work, collective if not collaborative in nature, and is resource intensive – i.e. the expensive work of infrastructuring. )

But such availability alone isn't enough for a more perverse interpretivist positivism – one harkening back. Here I agree with I much of what Andrew Moravcsik writes about availability -- not only data availability, but analytic availability as well:

Qualitative research's distinctive epistemology implies that to track the interpretation and analysis, a reader requires more than just access to a source. One must specify where within a descriptive or causal narrative each piece of evidence fits, and which specific textual passage in the source is critical. As historians, legal academics, and interpretivist social scientists insist, an informed reader needs to know not just what a scholar cites, but why. <sup>ii</sup>

In these kinds of systems, the fundamental units or structures are not data objects, then, but readings. The act of reading, or figuration, confounds interpretivism and positivism

Dissemination occurs in every data system, in every knowledge system. That's what they do. And even if I wouldn't regard those statements as "reproducible," meaning that you would expect some other empirical humanist to pore over and analyze the same material and weave into the same interpretive truth, I nevertheless regard them as true. And if you think interpretive is some kind of disqualifying qualifier that the real sciences have no need of—well, do you really think the situation isn't at least a little more perverse than the normally operative oppositions would suggest: real or interpreted, quantitative or qualitative, normal or perverse,

hard modernist positivist rationality or squishy postmodern depositivist bullshit? I know that it is, truly, a little more perverse than that, and rather than pathologizing that perversity and trying to purify it out of the system, we should be listening more carefully to it, having richer and more open-ended conversations within it, and even encouraging it

And I am keen to show my work, and my data, to anyone who wants to see it.

These perversities matter, I think, and don't think we don't understand the stakes. Air pollution regs, systematic assault on EPA expert bodies, etc. is now really the time to embark on an insistence of the complexity of truth, the indelibility of its normal perversions and perverse normalcies, the truth of deconstruction?

Confessing my perverse archival desires

What could be more perverse than to call for a new "unity of science" movement?

Positivism or interpretivism? The perverse reality is that each needs the other, relies on the other, is supplemented by the other interpretivism is not the other of positivism, but its perversion. Not nonoverlapping magisterial of knowledge, but one not two cultures, and not some bullshit dialectic third culture, but  $\sqrt{2}$  Cultures

I've been part of a collaborative effort to build digital infrastructure to better foreground and explore these paradoxes and contradictions that inhabit and indeed drive every complex analytic knowledge system, interpretive and/or positivist alike, and discuss this in more detail at different points below.

We have our doubts about it all – how could we not? Doubt is, after all, perversely integral to the truth-making practices that we, as cultural anthropologists, both document and study among environmental, health and data scientists, and are ourselves informed and traversed by. I try to articulate some of them here—fulfilling another part of what I think my charge was—while embedding those misgivings, properly if paradoxically, within our unshatterable commitments to data availability, transparency, and reliable and relevant truth—and their difficult perversities.

Positivism is anxious about non-reproducibility, it's most dreaded perversion. But this retroping gives insight I think: non-reproducibility is not a failure of positivist knowledge, but its ineradicable or irreducible limit. That perversion will always be with it, and that needs to be normalized and not pathologized. No amount of availability or transparency will ward it off. No matter how big you make Big Data, as a protective fetish against reproductive failure, it can't be big enough. Being a size queen is a fine and no doubt common perversion, but there's no need to normalize it and valorize it.

Take the perverse/normal couplet here:

Perversion and that inadequately specific term *normality* construct each other...[P]erversion and heteronormality constitute each other's limits. Perversion marks the boundary across

which you become an outlaw. Normality marks off the territory that, if stayed inside, keeps you safe from shame, disgust, and anxiety.” (Dimen 2001: 838)

And substitute the interpretivist/positivist couplet:

Interpretivism and that inadequately specific term *positivism* construct each other...Interpretivism and positivism constitute each other’s limits. Interpretivism marks the boundary across which you become an outlaw. Positivism marks off the territory that, if stayed inside, keeps you safe from shame, disgust, and anxiety.”

We also love, for example, [the “Shuffle” button placed at the top of every data-object in the ToxicDocs archive](#), and we are developing a similar function to introduce an element of chance into every encounter with ethnographic data, a bit of noise added back to every coded signal. Again, this is the kind of feature that has to be infra-structured in to one’s knowledge bases.

---

<sup>i</sup> “Infrastructures are matter that enable the movement of other matter. Their peculiar ontology lies in the fact that they are things and also the relation between things. As things they are present to the senses, yet they are also displaced in the focus on the matter they move around. We often see computers not cables, light not electricity, taps and water but not pipes and sewers...Yet the duality of infrastructures indicates that when they operate systemically they cannot be theorized in terms of the object alone. What distinguishes infrastructures from technologies is that they are objects that create the grounds on which other objects operate, and when they do so they operate as systems. Perhaps because of this duality, infrastructures are conceptually unruly.” Larkin, Brian. “The Politics and Poetics of Infrastructure.” *Annual Review of Anthropology* 42, no. 1 (2013): 327–43; <https://doi.org/10.1146/annurev-anthro-092412-155522>; p. 329.

<sup>ii</sup> Moravcsik, A. (2019). Transparency in Qualitative Research. In P. Atkinson, S. Delamont, A. Cernat, J.W. Sakshaug, & R.A. Williams (Eds.), *SAGE Research Methods Foundations*. doi: 10.4135/9781526421036863782