

## **Metadata, Digital Infrastructure, and the Data Ideologies of Cultural Anthropology**

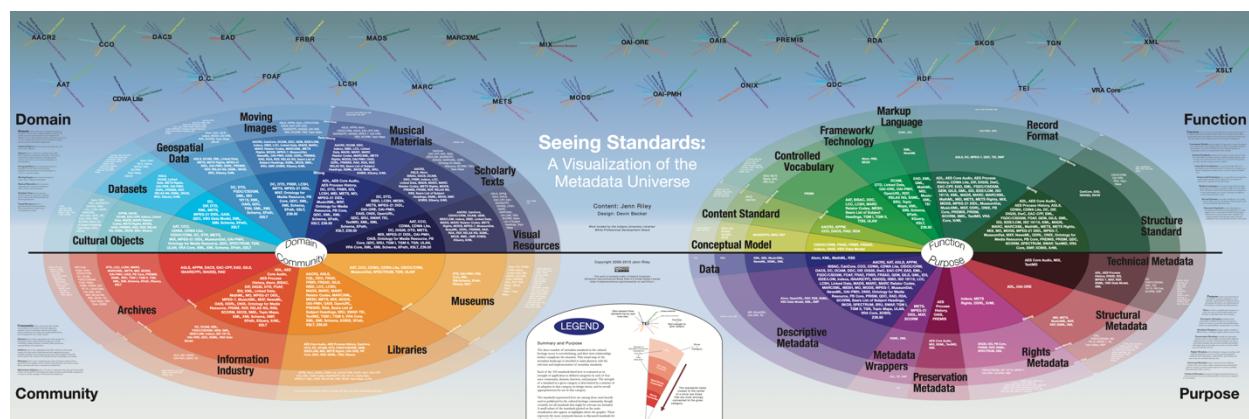
**Lindsay Poirier, Kim Fortun, Brandon Costelloe-Kuehn, and Mike Fortun**

(To be published late 2019 in *Anthropological Data in the Digital Age*, Jerome Crowder, Mike Fortun, Rachel Besara, and Lindsay Poirier, eds. London: Palgrave Macmillan.)

### **Introduction**

In 2010, digital librarian Jenn Riley (then at the Indiana University Libraries) created a comprehensive “glossary,” stunning in detail and scope, of the metadata standards most frequently deployed in the cultural heritage domain, from AACR2 to Z39.50, with 103 entries in between; Devin Becker designed the intricately structured visualization that accompanied the 18-page brochure. (See Figure 1.)

As broadly defined by Jane Greenberg, “metadata describes information and its context and associations; it is integral to the operation and function of any system preserving and supporting discovery, access, and use of information.” (Greenberg 2009:3610; see Smith et al, this volume). This universe of cultural heritage metadata standards was charted in two color-coded pinwheels, with the different colors representing the domain (Cultural Objects, Musical Materials, Scholarly Texts, etc.) in which a particular metadata standard is particularly strong (with the “strongest” standards closer to the center and the “weaker” ones out toward the edges). Arrayed above the two galactic clusters are the 30 most widely used standards displayed like variably-spiked stars, with the color and typographical weight of the domain name signifying the configuration of that particular metadata standard’s multiple strengths.



We use astronomical language to describe this visualization because the effect on us when we first encountered it was one well-known to humanities scholars, the effect of the sublime -- or as Riley herself later characterized it, the visualization is “overwhelming.” (Riley 2009-2010). As magnificent an achievement as Riley’s glossary and accompanying visualization are, it’s this dominant effect that most interests us here. We’ve seen a number of audiences at metadata-oriented workshops or similar events be overwhelmed by that image - not in a sense of wonder and excitement, but more in sense of awed terror. The experience of other “users”

may have varied, of course, but when we were working out our own metadata practices for the digital anthropology platform described below, we consulted this guide but never really used it in any pragmatic way; it mostly reinforced a sense of perplexity and hesitancy. With so many choices, it seemed highly likely that we'd make the wrong one.

The image depicts a situation that has plagued designers of data infrastructure for decades - that as more diverse research domains<sup>1</sup> are incorporated into the universe of open data, standards tend to proliferate, becoming less "standard" as they evolve to the specificities that diverse communities address with their data. While anthropologists might feel overwhelmed by this image, many designers of data infrastructure would smirk knowingly at it. For them, this is a problem that disrupts the possibility that data will be Findable, Accessible, Interoperable, and Reusable, or as they would summarize, "FAIR" (Wilkinson et al. 2016).

Writing as cultural anthropologists who study data infrastructures and their designers/builders, we have been working in and between these domains since at least 2012. When we look at this image, we see a complex sociotechnical system, one that has emerged from the heterogeneous values and modes of meaning-making that diverse researchers in diverse domains bring to their work--*must* bring to their work--of documenting and understanding complexity. To us, this image signifies the urgency of having anthropologists more engaged in more conversations around metadata and other data sharing infrastructure. Anthropologists, having long studied how meaning forms, operates, iterates, and shapes the ways people understand and act in their worlds, can make valuable contributions to these conversations, debates, and designs of data sharing infrastructure--including its own. Anthropology, we will argue, should be planning for a future of data preservation and sharing that doesn't depend on simply adopting technologies, models, best practices, and templates from other disciplines, other communities of practice, but instead enters into collaborative, creative interaction with the librarians, data scientists, and other information specialists engaged in developing such sociotechnical systems for anthropologists and beyond.

We use Riley's chart, then, not to reinforce a message that data work in anthropology is overwhelming in its complex demands, even though it can sometimes feel that way. We use it instead to illustrate anthropology's long-standing interests in, and long-cultivated capacities to make sense of, situations that can be overwhelming in their diverse and proliferating particularities and interactions across multiple scales. In the rest of this chapter we discuss some things we've learned from our experience in such situations as builders of the Platform for Experimental Collaborative Ethnography (PECE; <https://worldpece.org>) - a digital platform for archiving and sharing ethnographic data. Our work designing and developing this platform has drawn us into conversations around the future of research data sharing - both in anthropology, and in international and interdisciplinary organizations advancing open science. Our work on PECE has led us to become "observant participants" (Fassin and Rechtman 2009:11) in the Research Data Alliance (RDA; <https://rd-alliance.org>) - an international organization that aims to enable more open data sharing in diverse disciplines, although primarily in the natural sciences. Not only have both of these involvements given us valuable

fieldwork material on data and metadata practices and cultures, they have also positioned us as provocateurs. Building on growing support for open access publishing models, in discussions in cultural anthropology (at conferences, workshops, etc.), we have sought to call attention to the promises -- and real challenges -- of open data in cultural anthropology, both for democratizing access to anthropological data and for fostering ethical engagements with research interlocutors.

Through our collaborative engagements, we've become increasingly committed to the promise of future anthropologies in which more of our data are more openly available to be re-used and re-interpreted by other anthropologists, by researchers in other fields, and by diversifying data publics. Why are we so invested in sharing more data more openly in cultural anthropology -- or what in RDA we refer to more broadly as the "empirical humanities," including such related pursuits as folklore studies and oral history? As is the case with many of our fellow researchers in RDA, we are really only beginning to imagine why. While we acknowledge that there is much we still don't know about the promises and challenges of making qualitative data openly available for re-interpretation, we argue throughout this chapter that sharing data in cultural anthropology opens possibilities for making anthropology "thicker," more complex, and more enmeshed in the world's practicalities; for making anthropologists more collaborative, with each other and with many other others; and for enlisting a diverse readership, and for diversifying even further. To put it another way, and to introduce a term whose meanings we will elaborate in the next section: in many ways, we share a "data ideology" with our colleagues in RDA that sharing data, under appropriate circumstances, can advance better research, enliven collaboration, and bring about ethics and accountability in ethnographic practice.

We next elaborate on "data ideologies," and how we learned of their importance through our fieldwork within RDA and with the development of PECE. We'll also discuss how we found that, in order to further this promise of an anthropology that shares and re-uses more and more data, we need to better attend to our own data ideologies: to learn to articulate them more fully, to understand and operationalize them in materialities of digital infrastructure, and to narrate the emerging benefits we are starting to see, especially those coming from developing metadata practices that enable the re-use and re-interpretation of anthropological data.

### **What are data ideologies?**

In 2013, we found our way into the Research Data Alliance (RDA), initially through colleagues in computer science at Rensselaer Polytechnic Institute where a number of RDA-U.S. leaders are based. RDA is a global network (although concentrated in the U.S., Europe, and Australia) led by data, computer, and natural scientists, librarians, and similar researchers that are "building the social and technical bridges to enable open data sharing... across technologies, disciplines, and countries to address the grand challenges of society" (<https://www.rd-alliance.org/about-rda>). We joined RDA soon after it was formed in 2013, finding it a rich site for our varied fieldwork into how social and technical challenges to data sharing and digital

collaboration are imagined and tackled by a wide variety of technologists, data scientists, and “domain experts” including both natural and social sciences.<sup>2</sup> RDA leadership is acutely aware of the need to foreground the many social dimensions of interdisciplinary collaboration, and consistently brings its many conversations around data sharing back to issues of trust and culture (even if many within RDA would be hard-pressed to provide good articulations of these), and to the diverse and often conflicting understandings its different members have of even the most “technical” things (i.e. “protocols,” “ontologies,” “metadata”). Many in RDA have adopted a language of “sociotechnical” things and systems that was developed by anthropologists and others in the interdisciplinary field of Science and Technology Studies of which we are part. At their best, then, many of our interlocutors in RDA know that cultural work is always interlaced with technical work, and vice versa.

At the same time, we have observed within the RDA both dramatic limitations and variance in how the “social sciences and humanities” are generally understood (beginning with how they are so often grouped together), and how their data needs and practices should be conceptualized and managed. Roughly put, qualitative data and research are often cast within RDA as just slightly more complicated versions of quantitative data and research: both essentially positivist pursuits, guided by similar ideals of objectivity and reproducibility, and aimed at unified, singular explanations or truths, whether of the natural or sociocultural kind. The resulting need to repeatedly “explain ourselves” within RDA, and to articulate the differences in methods and theories of interpretive (hermeneutic) analysis on which cultural anthropologists (not to mention historians, urban studies scholars, folklorists, sociologists, ethnomethodologists, literary theorists, LGBTQ scholars, and sundry other scholars educated in diverse genealogies of the humanities and social sciences) rely has been a major, if not entirely planned, benefit of our engagement with this organization.

It was through this practice of explaining ourselves to researchers in RDA – and through observing other domain communities explain themselves to us and to each other – that we came recognize a need to examine and expose the diverse assumptions and commitments that different data sharing communities bring to their work. This is what has made RDA such a rich fieldsite for encountering and coming to understand its varied cultures of data sharing (and their difficulties). As we watch RDA members design standard ways to elicit feedback from domain communities, and as we hear them articulate “best practices” in data infrastructure design, we gain clues into how they (and we) perceive identity and difference, how they (and we) understand communication and collaboration to work, and how they (and we) value the production and preservation of knowledge. We observe how these perceptions, understandings, and values impact their (and our) deliberations and design choices, eventually interweaving into the digital architecture of data sharing infrastructure. We’ve come to understand how complex cultural and institutional forces shape particular, yet always collective, ideas and values about data sharing, as well as how these ideas shape various approaches to data infrastructure design, as constituting a community’s “data ideology.”

A data ideology is similar to, and perhaps even a version of, what linguistic anthropologists have identified as language ideology (Silverstein 1979; Woolard and Schiefflin 1994) or, more broadly, semiotic ideology -- what Webb Keane (2018:65) glosses as "people's underlying assumptions about what signs are, what functions signs do or do not serve, and what consequences they might or might not produce." Data are signs; they hold and carry meaning between agents in contexts; they point, index, signify, refer, represent, symbolize, accomplish, etc. Simply substituting one sign for another in Keane's formulation: we use "data ideology" to refer to, talk, and think about what people's underlying assumptions about data are, what functions data do or do not serve, and what consequences data might produce. Data ideologies are thus like a complex set of assumptions and understandings, both tacit and explicit, that form a kind of meta-discourse about data, how it functions, what needs to be done to and with it, who should handle it and how, and why it is valued -- and might be rendered still more valuable. It's worth stressing that a data or language ideology is not simply negative, something to be transcended or eliminated through reason and force of will, but is the wider space or set of language values shaping all cultural possibilities. As such, data ideologies have many dimensions and effects.

A growing literature has shown how incommensurate worldviews can impede consensus and produce "data friction" (Edwards et al. 2011) and how different communities might work to find "common ground" (Nafus 2017) -- that is, re-align, re-work, or re-define elements of their data ideologies. It has also unpacked the institutional and organizational barriers to advancing data sharing work -- citing issues such as rules that restrict agencies from sharing data with each other, reward structures that focus on individual achievement, and data privacy issues (Borgman 2012). In elaborating data ideology, we turn the kaleidoscope, offering different ethnographic takes on data sharing communities, aiming less to articulate what makes data sharing hard and more to simply affirm (while opening opportunities to critique) the diversity of values, commitments, and analytic thought styles that diverse researchers bring to their work in the often inchoate form of a data ideology.<sup>3</sup>

There are many data ideologies informing work at the RDA, constituted in a number of different cultural contexts or data domains, from anthropology itself to the data, environmental, genomic, and other sciences that our interlocutors practice. However, a few ideologies have become hegemonic in this space, orienting the types of topics discussed in plenary keynotes, who is considered to have the authority to raise and settle data infrastructure design disputes, and who is considered to have the expertise to write the standards that can cut across disciplinary (and ideological) differences.

For example, data scientists often refer to the acronym "FAIR" -- Findable, Accessible, Interoperable, and Reusable -- as a "gold standard" by which to evaluate data infrastructure and practice. (Even the metaphoric use of a term like "gold standard," which has had little purchase in real-world economic systems for decades if not much longer, suggests much about the data ideology and its work here.) Discourse advocating for researchers and repositories to make their data "FAIR" originated in the European Union in the mid-2010s,

and has more recently begun to gain traction in the United States. At the RDA, data scientists and scientific researchers often cite the concept of “FAIR”ness when referring to infrastructure needed to enable the discovery, legal retrieval, and use of others’ scientific data. The ‘I’ and the ‘R’ suggest that in order to reuse another scientists’ data, the data needs to be structured with open formats, described with widely used, machine-readable metadata, and licensed for re-use. The “FAIR”ness concept has spread rapidly in the data science community – some would even suggest earning buzzword status at the RDA and amongst other data sharing organizations.

When it comes to assessing the dominant data ideology at RDA, it is perhaps most important to point out topics that are not addressed when the community cites the acronym. On several occasions, we have heard data scientists pose the rhetorical question: “Who could argue with FAIRness?” – suggesting that data openness and reuse is considered a universal good. In doing so, they tend to depoliticize data sharing – ignoring the stakes and interests that some governmental actors, businesses, and even researchers and participants have for keeping data opaque and disaggregated. Further, as FAIRness was touted at the 12<sup>th</sup> RDA plenary held in Gaborone, Botswana (specifically to draw attention to data sharing challenges in parts of the world, like Botswana and many other nations in Africa, hitherto marginal to the data centers of U.S-European-Asian data worlds), neither large plenary nor smaller breakout sessions paid much attention to the colonial histories of extraction of both physical and intellectual property that have shaped the cultural and economic geography of the country (and the continent). In other words, the dominant data ideology tends to fetishize “sharing” – at times, eclipsing the complex ethico-political histories on which concepts such as transparency and ownership are founded.<sup>4</sup>

Furthermore, the urgency of designing “FAIR” data sharing infrastructure is often discussed, not in an ethico-political sense, but instead, in the context of science’s “reproducibility crisis” – a predicament emerging in scientific discourse suggesting that scientists often are unable to repeat a colleague’s experiment and reproduce the same statistically significant results. This has come as a hard blow to scientific communities, which have, since the 17th century, posited reproducibility as a fundamental component of the scientific method and the key ingredient in lending credibility to scientific claims. A dominant data ideology across RDA communities posits that the reproducibility crisis results not only from experiments being poorly designed, or from ever-rising institutional pressures to publish, but also a result of infrastructural shortcomings -- that researchers attempting to repeat an experiment based solely on information that they gather from published journal articles, cannot control for all the small judgments and decisions that the scientist that had originally conducted the study made along the way. If the data and workflows that serve as a basis for scientific experiments were open, accessible, and richly described – that is, if they were FAIR – other researchers could more readily recognize these judgments, test the results, and confirm the conclusions. In this context, pursuing scientific “truth,” first and foremost, demands (infra)structuring data so that there is uniformity in how scientists access, interpret, and use it.

Many researchers and practitioners working with the RDA acknowledge that sociocultural factors – such as the heterogeneity of languages and protocols scientific communities employ – pose barriers to naming, describing, organizing, and interpreting data consistently. Through working to enable interdisciplinary collaboration, they confront and are forced to recognize the diverse “epistemic cultures” (Knorr-Cetina 2009) that orient their work in different ways; however, culture is almost always cast as a problem to be overcome in these communities – something that limits or impedes progress towards scientific truth. To achieve reproducibility and uniformity, the dominant data ideology suggests that properly designed data infrastructure can and should control for the epistemological and semiotic differences amongst diverse scientific disciplines operating in diverse contexts. Notably, the ideology posits that these differences *must* be overcome in order to tackle the reproducibility crisis.

Few at the RDA would argue that overcoming these differences demands that all researchers adopt the same languages for naming and describing their data. Instead, data scientists at RDA advocate for the development of machine-readable thesauruses that map the relationships between terms used in different communities. According to the dominant data ideology, the heterogeneity of language across scientific disciplines can be reconciled by building translational layers *on top of* domain-specific vocabularies. Metadata are crucial to such efforts. We have often witnessed efforts to design “canonical” metadata standards that aim to federate persistently proliferating domain standards by identifying the most general terms needed to characterize all scientific research and mapping terms individual domain communities use onto these official terms. This makes compelling sense only by virtue of other features of the dominant data ideology: that there are one-to-one correspondences between the words used in different contexts, and that meanings can be mapped--directly, logically, without loss or shift of meaning--from one language system to the next.

Trying to understand and analyze the data ideologies within the RDA, and their dominant characteristics, has prompted us to pay attention to the data ideologies in play in our own discipline of cultural anthropology, while recognizing its diverse communities of practice. We’ve been moved by our RDA experience to consider how we might position our own efforts in the data infrastructure world to respect the various theoretical, methodological, and ethical commitments that cultural anthropologists bring to their work, while also encouraging them to think more expansively about what they can and should do with their (and with each other’s) research data. What have we learned about the data ideologies of cultural anthropology? We unpack this question in the following section.

### **Data Ideologies of Cultural Anthropology**

To our brief and partial sketch of the dominant data ideology of RDA we should add one note about its accompanying feel or style: there is an unmistakable and palpable sense of enthusiasm, excitement, and possibility in RDA culture(s) not conveyed by the measured and technocratic language of “FAIR”ness. Data is bound not only to such sober values as grounded stability, machinic transparency, and faithful reproducibility in this ideology; the

promise of sharing more data holds exciting and unknown potentials, extravagant possibilities not only for the future of one's own research and research field, but for new research collaborations and, indeed, for the larger world as a whole. Despite enormous sociotechnical challenges and difficulties, the mood at RDA plenaries is always upbeat and optimistic.

Enthusiasm and excitement are not the first words we would think of if asked to characterize cultural anthropology's own disciplinary culture when it comes to data and its sharing; caution, reticence, and worry are more fitting descriptors, in our experience, and outright resistance, hostility, or disinterested dismissal are sometimes part of the picture, too. When we've spoken about these matters, whether on conference panels or at colloquia or more informally with colleagues around these professional settings, there have been frequent demurrals and sometimes objections to our use of the word "data," as if it could never mobilize any meanings other than reductive and positivist ones, so inappropriate to the nuanced, subtle, irreducible complexities of people, their cultures, and their "worlding practices" (Roy 2011).

In this section we discuss some of these dimensions of our own disciplinary culture that have shaped a shared dominant data ideology concerning what anthropologists should or could do with the texts, photographs, audio and video recordings, and sundry digital and material artifacts we avidly collect and, in many ways, help co-create. How we think about and handle empirical materials in cultural anthropology is the product of a long disciplinary history, reflected in practices centered largely (and fittingly) on the protection of the people that cultural anthropologists have traditionally worked among, and their many vulnerabilities. Working to shift and multiply this data ideology in various ways will be challenging, demanding both cultural and technical development and working on the feedback loops to keep the whole enterprise moving.

We've learned a great deal about data cultures and data ideologies not only from our involvements with RDA, but also from our years of work in developing the Platform for Experimental Collaborative Ethnography (PECE) and the many worlds PECE has drawn us into over the years. PECE is, in brief, a Drupal-based digital platform designed to support collaborative and distributed interpretive analysis of ethnographic data while providing a general model for the archiving, sharing, and collaborative analysis of materials generated by empirical humanities scholars.

Using PECE in our own ethnographic research projects, and now helping new users set up new ethnographic data archiving and sharing projects of their own, has provided us with sharper insights into some of the particularities of cultural anthropology's data ideologies. Many of the data scientists and digital infrastructure builders in research domains from genomics to air chemistry that we've met in RDA have had to map their workflows, data types, and vocabularies in order to design and realize new data infrastructure. Similarly, in building PECE we have had to map the workflows of an anthropologist; to articulate the many "use cases" through which our typically more qualitative and less structured types of data move; to decide on and document vocabularies and protocols; and generally try to write, in digital form, how



“data ideologies” come to inhabit and inform anthropologists’ practice and thought. Our developing understanding of data ideologies in anthropology has largely emerged from introducing the PECE infrastructure (and the thinking and theorizing that has co-produced it) at a number of conferences, hosting and participating in formal and informal workshops,<sup>5</sup> collaborating with users of different PECE instances, and reading broadly in the literature on anthropology and data.

### *All or nothingism*

We begin our sketch of anthropology’s dominant data ideology with a story told by Robert Leopold, former director of the Smithsonian Institution’s National Anthropological Archives, while leading a group through a behind-the-scenes tour of some of the Smithsonian’s collections. Pausing in front of a display of Tlingit artifacts, he told the group about his recent delicate negotiations with a senior anthropologist who was resisting his plea to make her fieldnotes available to researchers sooner than fifty years hence, as she had requested. She believed that her notes were full of “culturally sensitive” information, such as the names of some people considered by others to be witches, and thus could not become open in any shorter time frame. “After much soul-searching and negotiation,” Leopold writes, “she and I devised a solution that I shared with my visitors”:

We would photocopy her original fieldnotes and redact the names of accused witches on a duplicate copy that we would provide to researchers, thereby allowing us to make the lion’s share of her field materials publicly accessible. I was sort of proud of my success.

At this point in my narrative, a student in the group spoke up: “I’m Tlingit,” she said. “Do you really think we don’t know if someone’s a witch?” (Leopold 2013: 86)

We can see a number of elements of a data ideology at work here, beginning with the unspoken assumption of many anthropologists that “fieldnotes” are a kind of totalized, one-size-fits-all category, for objects that are either too sacrosanct or too messily profane to be shared. When we talk enthusiastically about sharing field notes (a fairly common practice, in fact, among ethnographers who work in organizational or corporate ethnography), many if not most anthropologists are some combination of aghast and dismissive, and we get an immediate shy smile and a “oh, I could never share my fieldnotes,” as if they must be preserved perfectly intact and were all part of an inviolate set, with a varying but always noticeable emphasis on the extreme “never” and the possessive “my.” But once you say, “you know, we’re not talking about all of your fieldnotes; maybe you could just share some of them, after you’ve reviewed them and maybe edited them?,” the conversation has a better chance to get somewhere. Archivists like Leopold are generally more knowledgeable about the various practical options, often opened up by digital technologies for easily duplicating and manipulating objects, and for creating finer-grained distinctions about both categories of objects, and the levels and rules of privacy and accessibility that might attach to them.

### *Ethico-political Sensitivities*

A related dimension of a disciplinary data ideology widespread in anthropology is the presumption that all ethnographic data is by its very nature so sensitive, and one's interlocutors so vulnerable, that the best data posture to start from is a kind of protective crouch. There are of course good reasons for this, well-documented and with a disturbingly rich history. Neither Leopold nor we are suggesting that there is no such thing as culturally sensitive information, or that all anthropological data should simply be open. That, too, would partake of the same kind of all-or-nothing totalizing tendencies described above. But we do argue that *more* data can be made *more* shareable than anthropologists have traditionally held; that (often fairly simple) technical, social, and/or material practices can help that happen; and that included among the cultural obstacles to opening up anthropological data may be an anthropologist's own inflated sense of the special power held by our sorts of knowledge and the data from which it is crafted.

We'll stress again that there are indeed many situations (in our infrastructure development work we would call them "use cases") in cultural anthropological research where data can be sensitive, knowledge can harbor harm, and safeguards are the utmost concern. A data ideology with such elements embedded deeply within it is not only an understandable but a good thing. Ideologies, as cultural anthropologists should be the first to acknowledge, are not irrationalities to be overcome; they are spaces of meaning and action, the complexities of which we are continually challenged to understand more "thickly," and possibly to transform. Our broader point here is that the blanket protectionism in our dominant data ideology can too easily foreclose these kinds of assessments and self-assessments of power-effects. They may be inappropriately inflated, they may not; at the very least, they deserve reflection and analysis, and to be put into conversation with other perspectives, including of course the perspectives of those most intimately associated with the data: the people with whom they were generated.

In addition, it's long past time to recognize that, nearly fifty years after Laura Nader (1972) encouraged anthropologists to "study up" the gradients of power that cross our long-globalizing social worlds, our data and their ethico-political sensitivities are as diverse as the planet's people are--and as diverse as our ethnographic projects are. For the growing number of cultural anthropologists who work, often in close collaboration with people "in the field," to make sense of the worlds of high energy physics, high finance, NGO leaders and activists, government agencies of all types and at all levels, astronomy, climate science, and so many other contexts where the private/public polarities are reversed, or at least not so highly charged, a data ideology so tightly centered on and cathected to a one-dimensional and universalized standard of data privacy begins to lose its grip.

### *Institutionally-Reinforced Individualism*

This privacy-fixated, data-clutching ideology is renewed and reproduced through a number of logics and sensibilities. The individualism that has historically dominated anthropology's research practice and culture, and that of almost all humanities and social sciences research,

naturalizes possessive notions of “my data.” This is further reinforced by an academic administrative system and culture in which credit, reward, and advancement are almost exclusively individualized. We know well from our RDA experiences that our colleagues in the natural sciences also face the challenges of diversifying credit mechanisms to recognize collective data contributions, curation work, and involvement in multi-researcher and often multidisciplinary research collaborations more generally, but at least they have more extensive histories and practices of collaboration, and a broader culture (including patterns of funding) that is shifting more and more in response to complex problems often involving “coupled human-natural systems,” to validate and even valorize collaborative, transdisciplinary “team science.”<sup>6</sup> For all their emphasis on sociality and relationality, anthropologists (and many other qualitative social scientists and humanists) can be a pretty anti-social bunch, especially when it comes to sharing data; broader shifts in a research culture, provoked by increasingly urgent needs to collaborate within and across established disciplinary boundaries and cultures, should be understood and embraced as welcome and healthy developments for our field.

### *Interpretive Exceptionalism*

Our anthropological data ideologies have also taken shape around an interpretive exceptionalism, sometimes hyperinflated (understandably so) as a defensive response to the dominance of quantitative analysis, methods, and data models. This exceptionalism synergizes with our discipline’s traditional aspiration to “holistic understanding,” supposedly distinct from the reductionism to which the natural sciences are committed. In numerous meetings and informal discussions with anthropologists and qualitative researchers in neighboring fields, we’ve heard many expressions along the lines of “sharing my data doesn’t make any sense; my interpretations are intimately shaped by years spent with my [the possessive pronoun often shows up again here] interlocutors, the product of an exquisite attunement to the deep subtleties and unspoken nuances of lived and dynamic cultural complexities, the delicate interactive effect of a unique and irreproducible ‘human instrument’ immersed in extended fieldwork.” Constant comparisons to the natural sciences, whether made by university administrators or through our own self-comparisons, strengthen the dominance of the data ideology and its affective dimensions.

All too aware of our secondary status in the knowledge systems and hierarchies of the modern research university, we have ample good reason to emphasize at every opportunity our belief that anthropological knowledge is extremely powerful and highly sensitive and nuanced, that our qualitative data and interpretive methods provide us with richer, more situated, and more complex insights than the quantitative data and analyses of our esteemed colleagues in the biomedical and physical sciences. A defensiveness tends to creep in, and we stick up for a bullied anthropological tradition. But in protectively insisting on its virtuous qualities, real and valuable as those may be, we also reinforce a data ideology that works against data sharing and re-interpretations in the name of a kind of organic, holistic, somewhat ineffable and thus largely idiosyncratic relationship between fieldworker-interpreter and her/his data.

Here again, we are most concerned with the effects these broader disciplinary ideologies and contestations have on data ideologies, data sharing practices, and data infrastructures. Of course interview data, field notes, and interpretive analyses often can be extremely powerful and sensitive, and there are many situations in which questions of the openness and accessibility of anthropological data/analysis deserves the most scrupulous attention and the most stringent accession protocols. And we believe strongly that interpretation is unquestionably a creative analytic act, far more of an (organic) art than a (mechanical) science. But we also know, from our involvement with the diverse scientists in RDA, and from decades of scholarship in science studies, that even the most quantitative of data sets and analyses are layered with interpretive practices and demands, moments of abductive<sup>7</sup> reasoning or speculative insight, and creative surprise and serendipity. The differences between qualitative and quantitative data are more matters of degree than they are of kind; all researchers interpret.

In addition, the semiotic work of quantitative data scientists increasingly involves, and often directly focuses on, developing new data and metadata models and practices that capture or structure those interpretive dynamics and elements; anthropologists can learn from these threads of semiotic work and how they might extend our own understandings of what happens in “interpretation,” and how it might be further enriched and diversified. And finally, under different circumstances, these elements of interpretive exceptionalism constitute some of the very reasons why greater data sharing could, indeed should, be the default option in our data ideologies, rather than cause for a renewed valorization of a blanket, conservative protectionism underwritten by an ideology of holistic interpretive exceptionalism. Shouldn’t we want to maximize and leverage interpretive difference, rather than retreat behind it?

### **Beyond Reproducibility: The Re-interpretive Opportunities Metadata Affords**

Cultural anthropology’s commitments to interpretive practices also, however, shape its data ideology in more positive and productive directions. The data ideology dominant, indeed hegemonic, across virtually all the natural sciences and most of the social sciences, including the “qualitative” ones, is centered on the conception of reproducibility outlined above. The reasons why most researchers believe they should share data, why they should work so hard to make data FAIR, revolve around an epistemology of convergence and uniformity--the presumption that if you share enough data, and work hard enough and carefully enough on them using shared methods, there is one optimal solution on which everyone will ultimately agree. Epistemological pluralism or differences of interpretation are generally problems to be managed rather than resources to be tapped, the product of diverse “wild” cultures that need to be disciplined and tamed. That’s a powerful and necessary data ideology for many scientific pursuits, from drug design and clinical trials to social psychological studies, and an important driver of the “Open Science” movement.

Even efforts in more qualitative, cultural veins, like the long-running Human Relations Area Files (HRAF), are structured by these kinds of data ideologies for which interpretive difference is something to be transcended, or at least unified. In many ways, HRAF has since the 1950s

represented some of cultural anthropology's best impulses towards increased data sharing and collaborative, comparative analysis, facilitated through highly developed and standardized metadata practices. But the aims were a generalizability and a universalizable consensus; in the words of Melvin Ember, "the idea was to foster comparative research on humans in all their variety so that explanations of human behavior could escape being culture bound. An explanation that fits one society may not fit another" (Ember 2000:4). Systematically indexing ethnographic data according to HRAF's "Outline of Cultural Materials," with its over 700 subject categories, and making it available to researchers underwrote "the usefulness of social science," which in turn depended on the "validity of social science findings and theories."

If a finding or theory is not true under some circumstances, we would be foolhardy to think of applying it to real world situations and problems...A theory that seems to fit a particular region or even a sample of nations may not fit human societies generally...This is why HRAF was invented in the first place, to enable scientists to test their ideas about humans on worldwide data. (Ember 2000:6-7)

Productive as such a data ideology may be, it's not for all cultural anthropologists, many of whom know that their knowledge claims can be robust and trustworthy without participating in an ideology of reproducibility and monologic universality. While we may have a tradition of working to make sure (at least some of) our qualitative data is properly preserved (see Marsh and Punzalan 2019, this volume), our data ideology does not necessarily bind preservation to such one-dimensional notions of reproducibility, re-validation, hypothesis testing, and generalizability.

So while the FAIR commitment to a principle of data "re-use" is one cultural anthropology can share, these other elements of our data ideology position us to cast re-use in a more open-ended, pluralizing way. By situating data collection and analysis in a particular time, setting, and cultural context, metadata is key to encouraging re-interpretation of archived data, perpetually generating new, interpretive data which itself can be contextualized, archived, and re-interpreted. Jerome Crowder, for example, shows how re-analyzing his personally archived photographs from previous fieldwork in Bolivia through the technical metadata "uncover a rhythm of movement between the actors that is not apparent otherwise. The metadata expose the episodic nature of the work, our mutual engagement and my individual movements. Rather than compressing time, this ethnographic moment is understood in "real" time, inferring details about our relationship that were overlooked in my notes and dismissed by [my interlocutors] Basilia and Luis." (Crowder 2017:598). Such metadata-enabled re-interpretations of truly complex and multidimensional data allows us to access, analyze, and then re-analyze "the constant negotiation of positions, assumptions, and expectations that make up the intersubjectivity we share." (Ibid.:584) Sharing data ever more openly and widely is a way to deepen and diversify our understandings of our shared intersubjectivity and how it shapes, and can continue to reshape, anthropological knowledge.

Leveraging these differentiating, re-interpretive capacities and propensities towards openness that are also part of our disciplinary data ideologies can open up ethnographic data for still-to-be imagined purposes, with yet-to-be-named collaborators, not all of whom might be anthropologists. As one example, we point you to the Digital Repository of Ireland and their tremendous efforts to encourage, produce, and archive cultural data of many varieties in diverse projects, many in conjunction with the Irish Qualitative Data Archive. Here for example one can find full oral history transcripts, pseudonymized and with detailed transcripts including rich contextual metadata, of interviews with 23 women who worked in the infamous Magdalene Laundries, ready for re-interpretive uptake into any number of projects of classic anthropological and ethnographic interest: women's lives, reproductive rights, class structures, gender, religion, etc. (O'Donnell, Pembroke, and McGettrick 2015). The DRI/IQDA's new "Recall Initiative: from Memory and Life History, to Ireland and History" opens up the "archived reminiscences" of Irish women and men on topics from measles, mumps and scarlet fever to John F. Kennedy to Ireland's joining the European Union, in audio recordings open to analysis by their "interdisciplinary team of neuroscientists, historians, geographers, ethnographers and memory studies experts in order to explore the nature and construction of "cultural memory," and to trace the transition from autobiographical to semantic memory" -- open to anyone, for that matter, to listen to, to consider and reconsider, to re-interpret and re-figure (Allen 2018).

In a second example, in a special issue of *Sage Open* on "Digital Representations: Opportunities for Re-Using and Publishing Digital Qualitative Data," Florence Sutcliffe-Braithwaite shows us how interviews originally conducted to illuminate one subject (youth experiences of work and unemployment) can be reinterpreted for quite another (gender and sexuality) while helping us understand changing cultural discourses and strategies employed to "sidestep the dominant codes governing young, working-class women's sexuality." Sutcliffe-Braithwaite, a historian at University College London, was able to reanalyze the interview that her article centers around because, despite rather incomplete metadata<sup>8</sup> -- the interviewer is not even identified in the transcript! -- it was deposited at the University of Essex and is now included in the UK Data Archive. Her article uses reanalysis of empirically rich material to trouble "the ontology of 'prostitution' as a category," and we would argue that, along the way, she also troubles categories such as primary and secondary material. The figure and the ground (psychological forces, cultural contexts, interviewer and interviewee reflections) are constantly shifting in her article and this relatively fluid movement between "data" and "metadata" allows for incredibly rich analysis. In this example, more complete metadata (naming the interviewer, more "tags" instead of filing into discrete drawers<sup>9</sup>, and many of the metadata elements spelled out in Dublin Core, described below) could have radically accelerated Sutcliffe-Braithwaite's workflow and had myriad other benefits for her research process. Perhaps more importantly, for every intrepid Sutcliffe-Braithwaite willing to brave the limited context and imperfect metadata exemplified by this interview/archive, there are countless would-be reanalyzers that don't even attempt to take the first steps towards used shared data (or sharing their own) due, in part, to insufficient metadata and an insufficient imagination for what can be done with limited contextual information.

In addition to analyzing (and celebrating) many projects like the DRI/IADA and the reanalysis work of scholars like Sutcliffe-Braithwaite, our own work designing, developing, and deploying PECE in diverse contexts has assembled elements of anthropology's data ideologies (valuing nuanced interpretive analysis, situated/specific perspectives on data, the unique position of the anthropologist generating the data, etc.) in a way that explicitly aims to enable collaboration and experimentation, pushing back against some of the more anti-social and rigid tendencies common in the social sciences and humanities.

### **Re-interpretive Ideologies and Affordances in PECE**

We designed PECE as an open source (Drupal-based), virtual research environment to support collaboration among globally distributed researchers (primarily ethnographers) working with diverse data over extended periods of time. PECE was designed with qualitative data and theoretically-informed cultural analysis at its center. A signature feature supports the production and archiving of multiple interpretations of any given "artifact" (document, audio or video recording, image) -- what we have termed *collaborative hermeneutics*. PECE also archives the *structured analytics* (or sets of shared ethnographic questions) that stimulate interpretation, generating transparent workflows (which are usually individual and often remain tacit in qualitative research); archiving structured analytics also allows PECE users to easily move between different types and scales of analysis (using original and borrowed structured analytics), extending the dimensionality of their interpretations.

PECE was built within and for ethnographic projects, avoiding the problems associated with research infrastructure development that occurs without user involvement at the outset. To address IRB stipulations, the expectations of people studied, and researchers' own (varied) concerns, PECE offers stringent but flexible layers of privacy protection for data hosted in the platform. Data can be archived but fully restricted (accessible only to the researcher), partially restricted (accessible to IRB-approved collaborators), or openly, publicly accessible. We encourage all PECE adopters to build their projects around a presumption that as much data as possible be made available as openly and fully as possible. PECE also allows researchers to experiment with new ways of publishing their work to address diverse audiences, expanding multimodal anthropology through the incorporation of video or audio clips, images, and other media into one's text, while allowing for fieldnotes, working memos, collaborative annotations, and similar forms of "raw(er) data" to be presented as well (Fortun and Fortun 2019).

PECE also has clearly expressed *design logics* that convey the epistemic and aesthetic assumptions we built into the digital infrastructure of the platform, assumptions that reflect our commitments to a particular genealogy of work in cultural anthropology emerging in the late 1980s, in which we situate our own work as researchers and teachers (Fortun, Fortun, and Marcus 2017).

All this makes PECE an instance of what Johanna Drucker and Patrik Svensson call "middleware," their term for widely shared (digital) technical tools and infrastructures that

shape not only the collective experience and communications of users, but their patterns of thinking and interaction as well. Middleware platforms, in other words, from PowerPoint to Twitter to Wordpress to PECE, are never neutral; their design and material structures shape and color the inflected meanings they enable and convey. The humanities and humanistic social sciences need to focus and develop analytic attention to middleware of all sorts, Drucker and Svensson argue, both to how it functions and how it exceeds and/or fails those functions, to better understand the ways digital infrastructure indeed structures -- often invisibly -- its processes and products (Drucker and Svensson 2016; see also Dourish 2014; Goldberg 2015; see Franzen, this volume).

#### *What Metadata Affords in PECE*

As users upload ethnographic data into PECE as image, text, audio, or video artifacts, they are prompted to describe each file with the 15 core metadata elements defined in Dublin Core; this includes Title, Creator, Subject, Description, Publisher, Contributor, Date, Type, Format, Identifier, Source, Language, Relation, Coverage, and Rights. To us, the complex and conflicting data ideologies that informed the design of this metadata standard positions it to embody more than “recommended practice” in data management.

At the time of Dublin Core’s design in the mid-to-late 1990s, the information and computer scientists working to formalize the standard sought to define a set of “core” metadata elements that could be used to universally describe content on the newly emerging World Wide Web. Well-aware of the heterogeneity of “languages” different communities use to describe their data, they were worried about how metadata standards could proliferate and thus further splinter the way diverse communities represented their data. This became a key factor in debates that emerged between what the Dublin Core community referred to (in an acknowledged oversimplification) as the “structuralists” versus the “minimalists” (Weibel, Iannella, and Cathro 1997). Structuralists argued that the metadata standard should incorporate more complex structures<sup>10</sup> so that communities using different words or models to define and order their data could extend the standard to meet their specific needs. Minimalists, on the other hand, argued that the standard should be as simple as possible - only incorporating concepts for which the greatest number of people could agree upon a single, stable definition. Both groups wanted to prevent the proliferation of standards - the former, by adding structure to make Dublin Core more flexible to diverse needs, and the latter, by restricting the standard to what could be agreed upon as a core set of concepts.

While the standard morphed into a very minimalistic set of terms and definitions, the meanings encoded into Dublin Core have not been as singular or stable as either community would have liked. Some have even suggested that the metadata schema “died” around 2004, as the lack of precision in how its terms were defined, implemented, interpreted prevented the standard from performing the task it was originally designed for: enabling information seekers to retrieve the exact information they were seeking (Beall 2004).



For us, however, it is the very imprecision around definitions in Dublin Core that makes it such a robust standard for contextualizing our data. The metadata standard provides enough common structure to give anthropologists and other empirical humanists accessing and interpreting shared data some context as to where the data came from and who should be credited for its creation, contribution, and publication. However, we also acknowledge that, at least when it comes to anthropological data, setting canonical definitions for seemingly “standard” terms like “creator,” “rights,” and “subject” is a deeply political act, demanding that one takes a stance on what it means to create cultural narratives, what it means to claim ownership of those narratives, and what it means to impose a classification on them. In prompting users to fill out metadata profiles according to Dublin Core standards in PECE, we have sought to leverage the standard’s looseness and imprecision (design affordances that emerged from the conflicting data ideologies that constituted it) to trouble what it means to date, geographically pinpoint, credit, describe, or assign ownership to ethnographic data - all acts that were critiqued as part of anthropology’s increasing entanglement with poststructuralist, feminist, postcolonial, and other scholarship in the 1980s.

Wrapping ethnographic data with contextual information using this standard also enables us to build tools to extract files and their associated metadata from PECE and import them into other systems where the data can be reconfigured and reinterpreted. For instance, we have recently been in conversations with Ilya Zaslavsky at the University of San Diego about integrating PECE with a tool he built called Survey Analysis via Visual Exploration (SuAVE). SuAVE enables users to import structured survey data and images into a data exploration interface that offers functions for grouping, re-arranging, and visualizing data according to particular patterns characterized in the metadata. Because all image artifacts in PECE have been enriched with Dublin Core, we can extract the image files and their associated metadata from PECE and import them into SuAVE. In SuAVE, we can shuffle images by grouping and arranging them according to different patterns that emerge in the metadata - perhaps, viewing them chronologically, geographically, according to particular topics, or according to who has been credited. Notably, it is only because every image artifact in PECE has been described with the same set of metadata terms that we have the ability to define diverse groupings for the artifacts in SuAVE, which in turn enables us to remix the images and explore how ethnographic narratives shift as we view data through a kaleidoscopic lens.

Finally, in allowing users to create and archive shared sets of ethnographic questions, PECE is also designed to continuously generate new and evolving metadata around a particular artifact. As different researchers informed by different ethnographic traditions and genealogies respond to the same (evolving) questions to interpret shared ethnographic material, they can complement, extend and at times contradict collaborators’ interpretations, continuously deepening and differentiating the cultural narrative around ethnographic data. Rather than resolving how data should be defined, classified, or interpreted, these structured analytics elicit users to contribute new metadata in ways that leave space for open-ended, perpetually proliferating interpretations of data, which, in turn, highlights the constantly iterating nature of anthropological knowledge. Prompting diverse users to enrich the “meta”-narratives of a particular artifact, PECE’s structured analytics create opportunities to expose more than just the context of the data’s production; they also expose the context of data analysis. In this sense, responding to structured analytics highlights the provenance of thinking around ethnographic material, demystifying ethnographic workflows in ways that can open up anthropology - perhaps helping to usher more “civic anthropology” into the world.<sup>11</sup>

### **Metadata and the Semiotic Infrastructure to Come**

It will take a good deal of work and a good deal of time to shift our shared “data ideology” in the more open and pluralized directions briefly outlined above, directions we think better support the real analytic and cultural promise of shared anthropological data represented in projects like the “Recall Initiative.” We are also aware, good (enough) anthropologists that we are, that matters of data ideology are always intricately entwined with data materialities. This makes them resistant to easy or rapid change but also suggests those places -- in digital infrastructures, broadly speaking -- where changes can be identified, characterized, and effected. This is what we’re calling the semiotic work of digital infrastructure development, and central to it are new forms of metadata and new metadata practices to go with them.

The diverse, expanding, and somewhat disorienting metadata universe depicted by Riley at the opening of this chapter, therefore, is almost certain to become more diverse and more expansive but perhaps, paradoxically, less disorienting as anthropologists -- methodologically committed as we are to inviting excess and confusion, letting them illuminate and disrupt established habits and, over time, re-figuring new ones in shared intersubjective conversation with others -- become more adept through extended involvement in these new digital infrastructure and domains. Metadata models and practices that have been vital to the characterization and stabilizing preservation of (a small fraction of) cultural anthropological data will have to be extended, augmented, or otherwise re-formed or re-placed. We’ll need to invent new metadata forms and practices better suited to making our data more easily but justly shareable, more discoverable in wider and more diverse and even unknown use-contexts. Our kind of qualitative data and analyses -- special enough without needing to be extraordinarily or confoundingly so -- needs to be disseminated in new ways, among researchers familiar and strange, and between researchers and new publics. Our kind of qualitative data and analyses need to be re-interpreted in new ways, made available for re-use and re-analysis, more (but not necessarily fully) open so that it can be mashed up and meshed with other emergent data. It requires building new digital infrastructures, like the PECE and Mukurtu<sup>12</sup> platforms openly available on Github, to make new archival places open to more anthropologists rather than the few (elite, elder) beneficiaries of such limited sanctioned institutional data infrastructure as exists for conserving their data and knowledges.<sup>13</sup>

Every day, in numerous contexts, anthropologists old and young are generating new data and new analysis about human situations that, like memory in Ireland, are some mix of intellectually fascinating, politically vital, and culturally urgent, in locations from the most mundane and possibly imperiled to the most elite and developed but nonetheless constantly on the move, as all cultures are. These anthropologists might be involved in projects that try to understand and address “wicked” problems of sustainability, ecological and public health, democratic governance, and similarly urgent demands that transcend all disciplinary boundaries, and their data deserves and indeed needs to be more openly shareable than it is. Our infrastructure also needs to be more open to experimental tinkering with data and metadata models that capture just a bit more interpretive data about creative interpretive practice in all its fragmented, non-

holistic glory. All this requires linking infrastructural work (undervalued, underfunded, and underattended to in almost all fields, but especially so in anthropology) to the cultural work of shifting the discipline's data ideologies.

We'll therefore need to invent new forms of collaboration with other kinds of semiotic infrastructure experts: software developers, web and data scientists, and similar "technical" types. Work of this kind has to be experimental, not in an avant-gardist sense but in one we take from the sciences themselves: making careful perturbations to functioning systems to explore their latent capacities and affordances, guided by creative insight, observing and evaluating outcomes, and re-iterating the procedure. It's the kind of work we're still learning how to do through our continued hands-on development and design of the PECE platform -- a kind of sociotechnical work we've learned, anthropologically, in large part through our extended, distributed, engaged conversation with our collaborator friends in the RDA.

*Acknowledgements: the development of the PECE platform and the writing of this paper were made possible through the support of the National Science Foundation (Award #1535888), "Environmental Health Governance in Six Cities: How Scientific Cultures, Practices and Infrastructure Shape Governance Styles."*

## Notes

<sup>1</sup> David Ribes (2018:524) writes that "domains refer to those fields (often scientific, but not exclusively) concerned with worldly and specific matters, for example, linguistics is the "domain science" of language, biologists are the "domain experts" of organic life, and so on. The logic of domains parses the world into two main categories, one is either "in a domain" or one is working "independently" of any domain."

Typically, according to this "logic of domains," the computer, data, and information scientists designing data infrastructure for domain communities are considered to be working independently of any domain.

<sup>2</sup> Within the RDA, most of any group's work, including our own, occurs in the "spare time" of individual members in contact via electronic communications, but twice a year there is a large RDA "plenary" at which groups can meet face-to-face. We have used these plenaries (thirteen to date) in part as opportunities to convene sessions with other empirical humanists interested in advancing data infrastructure that is attuned to the specific challenges and needs of preserving and sharing such research. We have also attended the plenaries so that we can communicate our unique commitments and challenges to more technical groups attempting to develop data infrastructure that can facilitate data sharing across disciplinary borders.

<sup>3</sup> When querying a particular research community's data ideology, we consider questions such as:

- What does a particular research community seek to understand, and what kinds of data and analysis advance such understanding?
- How does a particular research community leverage theory and comparative perspective?
- What does a particular research community seek in collaboration?
- What does a particular research community seek to accomplish through their data representations, and what understandings of language, knowledge, and communication underpin their efforts?

<sup>4</sup> A reductive fetishization of sharing has structural parallels with at times naive mobilizations of "ecological" approaches, or a simplistic valorization of "connectedness" that can gloss over how

---

connections need not always be symbiotic and just; relationships can also be predatory, abusive, extractive, parasitic, etc.

<sup>5</sup> In addition to numerous colloquia at a number of universities, PECE has been presented, reviewed, and discussed as a collaborative opportunity at meetings held by: the American Anthropological Association, the Society for Cultural Anthropology, the Society for the Social Studies of Science, the Swiss Anthropological Association, the National Science Foundation, and the Research Data Alliance.

<sup>6</sup> In an early ethnographic moment in the prehistory of PECE, we attended an interdisciplinary conference on asthma, a complex example of “coupled human-natural systems” par excellence, and learned about just how difficult this interdisciplinary “team science” can be in practice. This ethnographic moment sparked our sense of the need for The Asthma Files (theasthmafiles.org), the first instance of PECE before it was formalized into a stand-alone digital infrastructure that can be downloaded and installed to support a range of ethnographic projects. In this case form (the platform) quite literally followed function (an assemblage enabling collaborative ethnography). The Asthma Files networks a wide variety of research and researchers all focused on asthma as a complex environmental health condition.

<sup>7</sup> Abduction was Charles Sanders Peirce’s term for a third mode of reasoning, a necessary companion to deduction and induction. It can be loosely translated to “hypothesizing” or, more loosely, “imaginatively guessing.” See Helmreich 2007 for a brief discussion in relation to anthropology.

<sup>8</sup> Sutcliffe-Braithwaite write that “it can sometimes be impossible to recover all the contextual information surrounding a particular interview. Yet it is still possible to reuse archived sociological data where not all the contextual information is available in the form we might want it.” We agree about this possibility, of course, and would argue that it is *always* impossible to recover all contextual information (or metadata) and that, as Derrida puts it in *Limited Inc.*, all communication and meaning exists only in context, and that that context can never be “saturated.” The idea that the context ever *could* be saturated, or fully “recovered” points to a particular language ideology that, we think, can be a barrier to more data sharing and iterations of analysis.

<sup>9</sup> The archive in which the interview was deposited in a collection called “Social and Political Implications of Household Work Strategies.” It was fortunate that Sutcliffe-Braithwaite had broad interests in labor and work, in addition to her focus on gender and sexuality, or she likely would not have encountered the interview she so deftly reanalyzes here.

<sup>10</sup> This division primarily manifested in debates over whether users should be able to “qualify” metadata fields - i.e. whether they should be able to attach additional attributes to metadata fields (metadata for metadata fields) to specify how they were defining/using that field in their own particular context. Minimalists argued that the core metadata elements should be as simple and consistent as possible; structuralists argued that indexers should be able to qualify these metadata elements.

<sup>11</sup> On “civic science” see Fortun and Fortun (2005)

<sup>12</sup> Mukurtu (mukurtu.org) is a digital platform similar to PECE that fulfills many of the same functions and aspirations to give communities a place to archive cultural heritage in diverse forms, and share it in ways they deem appropriate.

<sup>13</sup> In other words, we too want to make our data “FAIR” -- but, we also acknowledge that (like all metadata standards), as the concept of FAIRness begins to make waves in new research domains such as our own, its meaning will inevitably evolve. In the context of a cultural anthropological practice attuned to the re-interpretive possibilities afforded through metadata, FAIR may stand for Findable, Accessible, Interoperable, and *Reinterpretable*, may be guided by commitments to epistemological pluralism rather than reproducibility, and may signify the ethico-political sensibilities that anthropologists hope to advance through data sharing.

### Works Cited

Allen, A.P. (2018). Recall Initiative: From Memory and Life History, to Ireland and History. Digital Repository of Ireland [Distributor], Irish Qualitative Data Archive [Depositing Institution]. <https://doi.org/10.7486/DRI.wp98p078n>.

Beall, Jeffrey. (2004). Dublin Core: An Obituary. Library Hi Tech News 21(8): 40–41.

- 
- Borgman, Christine L. (2012.) The Conundrum of Sharing Research Data. *Journal of the American Society for Information Science and Technology* 63(6): 1059–1078.
- Edwards, Paul, Matthew S. Mayernik, Archer Batcheller, Geoffrey Bowker, and Christine Borgman. (2011) Science Friction: Data, Metadata, and Collaboration. *Social Studies of Science* 41(5): 667–690.
- Ember, Melvin. (2000). HRAF in the 21st Century: Blueprint for the Future After Fifty Years. Human Relations Area Files, Inc. New Haven, CT. <https://hraf.yale.edu/wp-content/uploads/2014/11/HRAF-in-the-21st-Century.pdf>
- Greenberg, Jane. (2009). Metadata and Digital Information. In *Encyclopedia of Library and Information Sciences*, Third Edition pp. 3610-3623. Taylor and Francis: New York. <http://dx.doi.org/10.1081/E-ELIS3-120044415>
- Helmreich, Stefan. (2007). “Induction, Deduction, Abduction, and the Logics of Race and Kinship.” *American Ethnologist* 34(2): 230–32. <https://doi.org/10.1525/ae.2007.34.2.230>.
- Fortun, Kim, and Mike Fortun. (2005). Scientific Imaginaries and Ethical Plateaus in Contemporary U.S. Toxicology. *American Anthropologist* 107(1): 43–54.
- Fortun, Mike, Kim Fortun, and George E. Marcus. (2017). Computers In/And Anthropology: The Poetics and Politics of Digitization. In *The Routledge Companion to Digital Ethnography*. Larissa Hjorth, Heather Horst, Anne Galloway, and Genevieve Bell, eds. Pp. 11–20. Taylor & Francis.
- Knorr-Cetina, Karin. (2009). *Epistemic Cultures: How the Sciences Make Knowledge*. Harvard University Press.
- Leopold, Robert. (2013). Articulating Culturally Sensitive Knowledge Online: A Cherokee Case Study. *Museum Anthropology Review* 7(1–2): 85–104.
- Marsh, Diana E., and Ricardo L. Punzalan. (2019). Studying and Mobilizing the Impacts of Anthropological Data in Archives. In *Anthropological Data in the Digital Age*. Jerome Crowder, Mike Fortun, Rachel Besara, and Lindsay Poirier, eds. Palgrave Macmillan.
- Nader, Laura. (1972). Up the Anthropologist: Perspectives Gained From Studying Up. <http://www.eric.ed.gov/ERICWebPortal/detail?accno=ED065375>, accessed October 12, 2012.
- Nafus, Dawn. (2017). Data Friction. Platypus: The CASTAC Blog. <http://blog.castac.org/2017/02/data-friction/>, accessed March 7, 2019.
- O’Donnell, Katherine, Sinead Pembroke, and Claire McGettrick. (2015). Magdalene Oral History Collection. Digital Repository of Ireland [Distributor], Irish Qualitative Data Archive [Depositing Institution]. <https://doi.org/10.7486/DRI.dn39x152w>.
- Ribes, David. (2019). STS, Meet Data Science, Once Again. *Science, Technology, & Human Values* 44(3): 514–539.
- Riley, Jenn. (2009-2010). Seeing Standards: A Visualization of the Metadata Universe. <http://jennriley.com/metadatamap/> Accessed March 2019.

---

Roy, Ananya. (2011). Urbanisms, Worlding Practices and the Theory of Planning. *Planning Theory* 10(1): 6–15.

Weibel, Stuart, Renato Iannella, and Warwick Cathro. (1997) The 4th Dublin Core Metadata Workshop Report. *D-Lib Magazine*, June.  
<http://mirror.dlib.org/dlib/june97/metadata/06weibel.html>.