

Metadata of the chapter that will be visualized online

Chapter Title	Metadata, Digital Infrastructure, and the Data Ideologies of Cultural Anthropology
Copyright Year	2020
Copyright Holder	The Author(s)
Corresponding Author	Family Name Poirier
	Particle
	Given Name Lindsay
	Suffix
	Division
	Organization/University University of California
	Address Davis, CA, USA
	Email lnpoirier@ucdavis.edu
Author	Family Name Fortun
	Particle
	Given Name Kim
	Suffix
	Division Department of Anthropology
	Organization/University University of California
	Address Davis, CA, USA
	Email kfortun@uci.edu
Author	Family Name Costelloe-Kuehn
	Particle
	Given Name Brandon
	Suffix
	Division
	Organization/University Rensselaer Polytechnic Institute
	Address Troy, NY, USA
Author	Family Name Fortun
	Particle
	Given Name Mike
	Suffix
	Division Department of Anthropology
	Organization/University University of California
	Address Davis, CA, USA
	Email fortunm@uci.edu

Abstract

In this chapter, based on fieldwork with the Research Data Alliance and our work designing the Platform for Experimental Collaborative Ethnography (PECE), we elaborate on the concept of data ideologies and examine how they have informed work and data-sharing practice in academic research, and in cultural anthropology more specifically. Data ideologies refer to people’s underlying assumptions about data, the way they operate, and the consequences they produce. We argue that, while many cultural anthropologists have been reticent to share their data, making anthropological data more open and accessible affords new possibilities for multi-perspectival analysis and re-interpretation of data—practices that can make ethnographic narratives more robust and pluralistic. Metadata is key to encouraging re-interpretation of archived data, as it situates data collection and analysis in a particular time, setting, and cultural context. We demonstrate how we implemented data-sharing infrastructure and metadata standards in PECE—not to advance reproducible research practices, but instead to encourage collaborative hermeneutics and iterative re-analysis of data. We conclude that attending to complex contemporary problems will demand linking undervalued and underfunded infrastructural work to the cultural work of shifting the discipline’s data ideologies.

Keywords (separated by “ - ”)

Metadata - Ethnography - Digital infrastructure - Collaboration - Re-interpretation - Data ideology - Hermeneutics

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

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

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 Fig. 10.1).

L. Poirier (✉)
University of California, Davis, CA, USA
e-mail: lnpoirier@ucdavis.edu

K. Fortun • M. Fortun
Department of Anthropology, University of California, Davis, CA, USA
e-mail: kfortun@uci.edu; fortunm@uci.edu

B. Costelloe-Kuehn
Rensselaer Polytechnic Institute, Troy, NY, USA



Fig. 10.1 Seeing standards: A visualization of the metadata universe

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

27 We use astronomical language to describe this visualization because the
 28 effect on us when we first encountered it was one well known to humani-
 29 ties scholars, the effect of the sublime—or as Riley herself later character-
 30 ized it, the visualization is “overwhelming” (Riley 2009–2010). As
 31 magnificent an achievement as Riley’s glossary and accompanying visual-
 32 ization are, it’s this dominant effect that most interests us here. We’ve seen
 33 a number of audiences at metadata-oriented workshops or similar events
 34 be overwhelmed by that image—not in a sense of wonder and excitement,
 35 but more in sense of awed terror. The experience of other “users” may
 36 have varied, of course, but when we were working out our own metadata
 37 practices for the digital anthropology platform described below, we con-
 38 sulted this guide but never really used it in any pragmatic way; it mostly
 39 reinforced a sense of perplexity and hesitancy. With so many choices, it
 40 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 domains¹ 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

80 research data sharing—both in anthropology, and in international and
81 interdisciplinary organizations advancing open science. Our work on
82 PECE has led us to become “observant participants” (Fassin and Rechtman
83 2009: 11) in the Research Data Alliance (RDA; [https://rd-alliance.
84 org](https://rd-alliance.org))—an international organization that aims to enable more open data
85 sharing in diverse disciplines, although primarily in the natural sciences.
86 Not only have both of these involvements given us valuable fieldwork
87 material on data and metadata practices and cultures; they have also posi-
88 tioned us as provocateurs. Building on growing support for open-access
89 publishing models, in discussions in cultural anthropology (at confer-
90 ences, workshops, etc.), we have sought to call attention to the prom-
91 ises—and real challenges—of open data in cultural anthropology, both for
92 democratizing access to anthropological data and for fostering ethical
93 engagements with research interlocutors.

94 Through our collaborative engagements, we’ve become increasingly
95 committed to the promise of future anthropologies in which more of our
96 data are more openly available to be re-used and re-interpreted by other
97 anthropologists, by researchers in other fields, and by diversifying data
98 publics. Why are we so invested in sharing more data more openly in cul-
99 tural anthropology—or what in RDA we refer to more broadly as the
100 “empirical humanities,” including such related pursuits as folklore studies
101 and oral history? As is the case with many of our fellow researchers in
102 RDA, we are really only beginning to imagine why. While we acknowledge
103 that there is much we still don’t know about the promises and challenges
104 of making qualitative data openly available for re-interpretation, we argue
105 throughout this chapter that sharing data in cultural anthropology opens
106 possibilities for making anthropology “thicker,” more complex, and more
107 enmeshed in the world’s practicalities; for making anthropologists more
108 collaborative, with each other and with many other others; for enlisting a
109 diverse readership; and for diversifying even further. To put it another way,
110 and to introduce a term whose meanings we will elaborate in the next sec-
111 tion: in many ways, we share a “data ideology” with our colleagues in
112 RDA that sharing data, under appropriate circumstances, can advance bet-
113 ter research, enliven collaboration, and bring about ethics and account-
114 ability in ethnographic practice.

115 We next elaborate on “data ideologies,” and how we learned of their
116 importance through our fieldwork within RDA and with the development
117 of PECE. We’ll also discuss how we found that, in order to further this
118 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.² 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 (often articulated in a functionalist sense), 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 (STS) 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

156 ideals of objectivity and reproducibility, and aimed at unified, singular
157 explanations or truths, whether of the natural or sociocultural kind. The
158 resulting need to repeatedly “explain ourselves” within RDA, and to
159 articulate the differences in methods and theories of interpretive (herme-
160 neutic) analysis on which cultural anthropologists (not to mention histo-
161 rians, urban studies scholars, folklorists, sociologists, ethnomethodologists,
162 literary theorists, LGBTQ [lesbian, gay, bisexual, transgender and queer
163 or questioning] scholars, and sundry other scholars educated in diverse
164 genealogies of the humanities and social sciences) rely has been a major, if
165 not entirely planned, benefit of our engagement with this organization.

166 It was through this practice of explaining ourselves to researchers in
167 RDA—and through observing other domain communities explain them-
168 selves to us and to each other—that we came to recognize a need to exam-
169 ine and expose the diverse assumptions and commitments that different
170 data-sharing communities bring to their work. This is what has made RDA
171 such a rich field site for encountering and coming to understand its varied
172 cultures of data sharing (and their difficulties). As we have watched RDA
173 members design standard ways to elicit feedback from domain communi-
174 ties, and as we have heard them articulate “best practices” in data infra-
175 structure design, we have gained clues into how they (and we) perceive
176 identity and difference, how they (and we) understand communication
177 and collaboration to work, and how they (and we) value the production
178 and preservation of knowledge. We have observed how these perceptions,
179 understandings, and values impact their (and our) deliberations and design
180 choices, eventually interweaving into the digital architecture of data-
181 sharing infrastructure. In the process, we’ve come to understand a com-
182 munity’s “data ideology” as constituted through the complex cultural and
183 institutional forces that shape particular, yet always collective, ideas and
184 values about data sharing and data infrastructure design.

185 A data ideology is similar to, and perhaps even a version of, what lin-
186 guistic anthropologists have identified as language ideology (Silverstein
187 1979; Woolard and Schieffelin 1994) or, more broadly, semiotic ideol-
188 ogy—what Webb Keane (2018: 65) describes as “people’s underlying
189 assumptions about what signs are, what functions signs do or do not serve,
190 and what consequences they might or might not produce.” Data are signs;
191 they hold and carry meaning between agents in contexts; they point,
192 index, signify, refer, represent, symbolize, accomplish, and so on. Simply
193 substituting one sign for another in Keane’s formulation: we use “data
194 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 a complex set of assumptions and understandings, both tacit and explicit, that form a 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.³

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 (Wilkinson et al. 2016). (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.)

234 Discourse advocating for researchers and repositories to make their data
235 “FAIR” originated in the European Union in the mid-2010s, and has
236 more recently begun to gain traction in the U.S. At the RDA, data
237 scientists and scientific researchers often cite the concept of “FAIR”ness
238 when referring to infrastructure needed to enable the discovery, legal
239 retrieval, and use of others’ scientific data. The “I” and the “R” suggest
240 that in order to re-use another scientists’ data, the data needs to be struc-
241 tured with open formats, described with widely used, machine-readable
242 metadata, and licensed for re-use. The “FAIR”ness concept has spread
243 rapidly in the data science community—some would even suggest earning
244 buzzword status at the RDA and among other data-sharing organizations.

245 When it comes to assessing the dominant data ideology at RDA, it is
246 perhaps most important to point out topics that are not addressed when
247 the community cites the acronym. On several occasions, we have heard
248 data scientists pose the rhetorical question: “Who could argue with
249 FAIRness?”—suggesting that data openness and re-use is considered a
250 universal good. In doing so, they tend to depoliticize data sharing—ignor-
251 ing the stakes and interests that some governmental actors, businesses, and
252 even researchers and participants have for keeping data opaque and disag-
253 gregated. Further, as FAIRness was touted at the 12th RDA plenary held
254 in Gaborone, Botswana (specifically to draw attention to data-sharing
255 challenges in parts of the world like Botswana and many other nations in
256 Africa, hitherto marginal to the data centers of U.S.-European-Asian data
257 worlds), neither large plenary nor smaller breakout sessions paid much
258 attention to the colonial histories of extraction of both physical and intel-
259 lectual property that have shaped the cultural and economic geography of
260 the country and the continent. In other words, the dominant data ideol-
261 ogy tends to fetishize “sharing”—at times, eclipsing the complex ethico-
262 political histories on which concepts such as transparency and ownership
263 are founded.⁴

264 Furthermore, the urgency of designing “FAIR” data-sharing infra-
265 structure is often discussed, not in an ethico-political sense, but instead in
266 the context of science’s “reproducibility crisis”—a predicament emerging
267 in scientific discourse suggesting that scientists often are unable to repeat
268 a colleague’s experiment and reproduce the same statistically significant
269 results. This has come as a hard blow to scientific communities, which
270 have, since the seventeenth century, posited reproducibility as a funda-
271 mental component of the scientific method and the key ingredient in lend-
272 ing 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 toward 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 among 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

312 compelling sense only by virtue of other beliefs constituting the dominant
313 data ideology: that there are one-to-one correspondences between the
314 words used in different contexts, and that meanings can be mapped—
315 directly, logically, without loss or shift of meaning—from one language
316 system to the next.

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

328 DATA IDEOLOGIES OF CULTURAL ANTHROPOLOGY

329 To our brief and partial sketch of the dominant data ideology of RDA, we
330 should add one note about its accompanying feel or style: there is an
331 unmistakable and palpable sense of enthusiasm, excitement, and possibil-
332 ity in RDA culture(s) not conveyed by the measured and technocratic
333 language of "FAIR"ness. Data is bound not only to such sober values as
334 grounded stability, machinic transparency, and faithful reproducibility in
335 this ideology; the promise of sharing more data holds exciting and
336 unknown potentials, extravagant possibilities not only for the future of
337 one's own research and research field, but for new research collaborations
338 and, indeed, for the larger world as a whole. Despite enormous sociotech-
339 nical challenges and difficulties, the mood at RDA plenaries is always
340 upbeat and optimistic.

341 Enthusiasm and excitement are not the first words we would think of if
342 asked to characterize cultural anthropology's own disciplinary culture
343 when it comes to data and its sharing; caution, reticence, and worry are
344 more fitting descriptors, in our experience, and outright resistance, hostil-
345 ity, or disinterested dismissal are sometimes part of the picture, too. When
346 we've spoken about these matters, whether on conference panels or at
347 colloquia or more informally with colleagues around these professional
348 settings, there have been frequent demurrals and sometimes objections to

our use of the word “data,” as if it could never mobilize any meanings 349
 other than reductive and positivist ones, so inappropriate to the nuanced, 350
 subtle, irreducible complexities of people, their cultures, and their “world- 351
 ing practices” (Roy 2011). 352

In this section we discuss some of these dimensions of our own disci- 353
 plinary culture that have shaped a shared dominant data ideology con- 354
 cerning what anthropologists should or could do with the texts, 355
 photographs, audio and video recordings, and sundry digital and material 356
 artifacts we avidly collect and, in many ways, help co-create. How we think 357
 about and handle empirical materials in cultural anthropology is the prod- 358
 uct of a long disciplinary history, reflected in practices centered largely 359
 (and fittingly) on the protection of the people that cultural anthropolo- 360
 gists have traditionally worked among, and their many vulnerabilities. 361
 Working to shift and multiply this data ideology in various ways will be 362
 challenging, demanding both cultural and technical development and 363
 working on the feedback loops to keep the whole enterprise moving. 364

We’ve learned a great deal about data cultures and data ideologies not 365
 only from our involvements with RDA, but also from our years of work in 366
 developing the Platform for Experimental Collaborative Ethnography 367
 (PECE) and the many worlds PECE has drawn us into over the years. 368
 PECE is, in brief, a Drupal-based digital platform designed to support 369
 collaborative and distributed interpretive analysis of ethnographic data 370
 while providing a general model for the archiving, sharing, and collabora- 371
 tive analysis of materials generated by empirical humanities scholars. 372

Using PECE in our own ethnographic research projects, and now help- 373
 ing new users set up new ethnographic data archiving and sharing projects 374
 of their own, has provided us with sharper insights into some of the par- 375
 ticularities of cultural anthropology’s data ideologies. Many of the data 376
 scientists and digital infrastructure builders in research domains from 377
 genomics to air chemistry that we’ve met in RDA have had to map their 378
 workflows, data types, and vocabularies in order to design and realize new 379
 data infrastructure. Similarly, in building PECE, we have had to map the 380
 workflows of an anthropologist; to articulate the many “use cases” through 381
 which our typically more qualitative and less structured types of data 382
 move; to decide on and document vocabularies and protocols; and gener- 383
 ally try to write, in digital form, how “data ideologies” come to inhabit 384
 and inform anthropologists’ practice and thought. Our developing under- 385
 standing of data ideologies in anthropology has largely emerged from 386
 introducing the PECE infrastructure (and the thinking and theorizing 387

388 that has co-produced it) at a number of conferences, hosting and partici-
 389 pating in formal and informal workshops,⁵ collaborating with users of
 390 different PECE instances, and reading broadly in the literature on anthro-
 391 pology and data.

392

All or Nothingism

393 We begin our sketch of anthropology's dominant data ideology with a
 394 story told by Robert Leopold, former director of the Smithsonian
 395 Institution's National Anthropological Archives, while leading a group
 396 through a behind-the-scenes tour of some of the Smithsonian's collec-
 397 tions. Pausing in front of a display of Tlingit artifacts, he told the group
 398 about his recent delicate negotiations with a senior anthropologist who
 399 was resisting his plea to make her field notes available to researchers sooner
 400 than 50 years hence, as she had requested. She believed that her notes
 401 were full of "culturally sensitive" information, such as the names of some
 402 people considered by others to be witches, and thus could not become
 403 open in any shorter time frame. "After much soul-searching and negotia-
 404 tion," Leopold writes, "she and I devised a solution that I shared with my
 405 visitors":

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

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

413 We can see a number of elements of a data ideology at work here,
 414 beginning with the unspoken assumption of many anthropologists that
 415 "fieldnotes" are a kind of totalized, one-size-fits-all category, for objects
 416 that are either too sacrosanct or too messily profane to be shared. When
 417 we talk enthusiastically about sharing field notes (a fairly common prac-
 418 tice, in fact, among ethnographers who work in organizational or corpo-
 419 rate ethnography), many if not most anthropologists are some combination
 420 of aghast and dismissive, and we get an immediate shy smile and a "oh,
 421 I could never share my fieldnotes," as if they must be preserved per-
 422 fectly 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 423
 once you say, “you know, we’re not talking about all of your fieldnotes; 424
 maybe you could just share some of them, after you’ve reviewed them and 425
 maybe edited them?,” the conversation has a better chance to get some- 426
 where. Archivists like Leopold are generally more knowledgeable about 427
 the various practical options, often opened up by digital technologies for 428
 easily duplicating and manipulating objects, and for creating finer-grained 429
 distinctions about both categories of objects, and the levels and rules of 430
 privacy and accessibility that might attach to them. 431

Ethico-political Sensitivities 432

A related dimension of a disciplinary data ideology widespread in anthro- 433
 pology is the presumption that all ethnographic data is by its very nature 434
 so sensitive, and one’s interlocutors so vulnerable, that the best data pos- 435
 ture to start from is a kind of protective crouch. There are of course good 436
 reasons for this, well-documented and with a disturbingly rich history. 437
 Neither Leopold nor we are suggesting that there is no such thing as cul- 438
 turally sensitive information, or that all anthropological data should simply 439
 be open. That, too, would partake of the same kind of all-or-nothing 440
 totalizing tendencies described above. But we do argue that *more* data can 441
 be made *more* shareable than anthropologists have traditionally held; that 442
 (often fairly simple) technical, social, and/or material practices can help 443
 that happen; and that included among the cultural obstacles to opening 444
 up anthropological data may be an anthropologist’s own inflated sense of 445
 the special power held by our sorts of knowledge and the data from which 446
 it is crafted. 447

We’ll stress again that there are indeed many situations (in our infra- 448
 structure development work we would call them “use cases”) in cultural 449
 anthropological research where data can be sensitive, knowledge can har- 450
 bor harm, and safeguards are the utmost concern. A data ideology with 451
 such elements embedded deeply within it is not only an understandable 452
 but a good thing. Ideologies, as cultural anthropologists should be the 453
 first to acknowledge, are not irrationalities to be overcome; they are spaces 454
 of meaning and action, the complexities of which we are continually chal- 455
 lenged to understand more “thickly” and possibly to transform. Our 456
 broader point here is that the blanket protectionism in our dominant data 457
 ideology can too easily foreclose these kinds of assessments and self- 458
 assessments of power-effects. They may be inappropriately inflated, or 459

460 they may not; at the very least, they deserve reflection and analysis, and to
 461 be put into conversation with other perspectives, including, of course, the
 462 perspectives of those most intimately associated with the data: the people
 463 with whom they were generated.

464 In addition, it's long past time to recognize that, nearly 50 years after
 465 Laura Nader (1972) encouraged anthropologists to "study up," the gra-
 466 dients of power that cross our long-globalizing social worlds, our data,
 467 and their ethico-political sensitivities are as diverse as the planet's people
 468 are—and as diverse as our ethnographic projects are. For the growing
 469 number of cultural anthropologists who work, often in close collaboration
 470 with people "in the field," to make sense of the worlds of high energy
 471 physics, high finance, non-governmental organization (NGO) leaders and
 472 activists, government agencies of all types and at all levels, astronomy, cli-
 473 mate science, and so many other contexts where the private/public polar-
 474 ities are reversed, or at least not so highly charged, a data ideology so
 475 tightly centered on and cathected to a one-dimensional and universalized
 476 standard of data privacy begins to lose its grip.

477 *Institutionally Reinforced Individualism*

478 This privacy-fixated, data-clutching ideology is renewed and reproduced
 479 through a number of logics and sensibilities. The individualism that has
 480 historically dominated anthropology's research practice and culture, and
 481 that of almost all humanities and social sciences research, naturalizes pos-
 482 sessive notions of "my data." This is further reinforced by an academic
 483 administrative system and culture in which credit, reward, and advance-
 484 ment are almost exclusively individualized. We know well from our RDA
 485 experiences that our colleagues in the natural sciences also face the chal-
 486 lenges of diversifying credit mechanisms to recognize collective data con-
 487 tributions, curation work, and involvement in multi-researcher and often
 488 multidisciplinary research collaborations more generally, but at least they
 489 have more extensive histories and practices of collaboration, and a broader
 490 culture (including patterns of funding) that is shifting more and more in
 491 response to complex problems often involving "coupled human-natural
 492 systems," to validate and even valorize collaborative, transdisciplinary
 493 "team science."⁶ For all their emphasis on sociality and relationality,
 494 anthropologists (and many other qualitative social scientists and human-
 495 ists) can be a pretty anti-social bunch, especially when it comes to sharing
 496 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 500

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 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.

532 Here again, we are most concerned with the effects these broader dis-
 533 ciplinary ideologies and contestations have on data ideologies, data-
 534 sharing practices, and data infrastructures. Of course interview data, field
 535 notes, and interpretive analyses often can be extremely powerful and
 536 sensitive, and there are many situations in which questions of the openness
 537 and accessibility of anthropological data/analysis deserves the most scru-
 538 pulous attention and the most stringent accession protocols. And we
 539 believe strongly that interpretation is unquestionably a creative analytic
 540 act, far more of an (organic) art than a (mechanical) science. But we also
 541 know, from our involvement with the diverse scientists in RDA, and from
 542 decades of scholarship in science studies, that even the most quantitative
 543 of data sets and analyses are layered with interpretive practices and
 544 demands, moments of abductive⁷ reasoning or speculative insight, and
 545 creative surprise and serendipity. The differences between qualitative and
 546 quantitative data are more matters of degree than they are of kind; all
 547 researchers interpret.

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

561 BEYOND REPRODUCIBILITY: THE RE-INTERPRETIVE 562 OPPORTUNITIES METADATA AFFORDS

563 Cultural anthropology’s commitments to interpretive practices also, how-
 564 ever, shape its data ideology in more positive and productive directions.
 565 The data ideology dominant, indeed hegemonic, across virtually all the
 566 natural sciences and most of the social sciences, including the “qualitative”
 567 ones, is centered on the conception of reproducibility outlined above. The
 568 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 toward 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

607 necessarily bind preservation to such one-dimensional notions of repro-
608 ducibility, re-validation, hypothesis testing, and generalizability.

609 So while the FAIR commitment to a principle of data “re-use” is one
610 cultural anthropology can share, these other elements of our data ideology
611 position us to cast re-use in a more open-ended, pluralizing way. By situat-
612 ing data collection and analysis in a particular time, setting, and cultural
613 context, metadata is key to encouraging re-interpretation of archived data,
614 perpetually generating new, interpretive data which itself can be contextu-
615 alized, archived, and re-interpreted. Jerome Crowder, for example, shows
616 how re-analyzing his personally archived photographs from previous field-
617 work in Bolivia through the technical metadata “uncover a rhythm of
618 movement between the actors that is not apparent otherwise. The meta-
619 data expose the episodic nature of the work, our mutual engagement, and
620 my individual movements. Rather than compressing time, this ethno-
621 graphic moment is understood in ‘real’ time, inferring details about our
622 relationship that were overlooked in my notes and dismissed by [my inter-
623 locutors] Basilia and Luis” (Crowder 2017: 598). Such metadata-enabled
624 re-interpretations of truly complex and multidimensional data allow us to
625 access, analyze, and then re-analyze “the constant negotiation of posi-
626 tions, assumptions, and expectations that make up the intersubjectivity we
627 share” (Ibid.: 584). Sharing data ever more openly and widely is a way to
628 deepen and diversify our understandings of our shared intersubjectivity
629 and how it shapes, and can continue to reshape, anthropological
630 knowledge.

631 Leveraging these differentiating, re-interpretive capacities and propen-
632 sities toward openness that are also part of our disciplinary data ideologies
633 can open up ethnographic data for still-to-be imagined purposes, with
634 yet-to-be-named collaborators, not all of whom might be anthropologists.
635 As one example, we point you to the Digital Repository of Ireland and
636 their tremendous efforts to encourage, produce, and archive cultural data
637 of many varieties in diverse projects, many in conjunction with the Irish
638 Qualitative Data Archive. Here for example one can find full oral history
639 transcripts, pseudonymized and with detailed transcripts including rich
640 contextual metadata, of interviews with 23 women who worked in the
641 infamous Magdalene Laundries, ready for re-interpretive uptake into any
642 number of projects of classic anthropological and ethnographic interest:
643 women’s lives, reproductive rights, class structures, gender, religion, and
644 so on (O’Donnell et al. 2015). The DRI/IQDA’s new “Recall Initiative:
645 from Memory and Life History, to Ireland and History” opens up the
646 “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 re-interpreted 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 re-analyze the interview that her article centers around because, despite rather incomplete metadata⁸ (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 re-analysis 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,⁹ 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 re-analysts that don’t even attempt to take the first steps toward 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 re-analysis 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

687 nuanced interpretive analysis, situated/specific perspectives on data, the
688 unique position of the anthropologist generating the data, etc.) in a way
689 that explicitly aims to enable collaboration and experimentation, pushing
690 back against some of the more anti-social and rigid tendencies common in
691 the social sciences and humanities.

692 RE-INTERPRETIVE IDEOLOGIES AND AFFORDANCES IN PECE

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

707 PECE was built within and for ethnographic projects, avoiding the
708 problems associated with research infrastructure development that occurs
709 without user involvement at the outset.

710 To address IRB stipulations, the expectations of people studied, and
711 researchers’ own (varied) concerns, PECE offers stringent but flexible lay-
712 ers of privacy protection for data hosted in the platform. Data can be
713 archived but fully restricted (accessible only to the researcher), partially
714 restricted (accessible to IRB-approved collaborators), or openly, publicly
715 accessible. We encourage all PECE adopters to build their projects around
716 a presumption that as much data as possible be made available as openly
717 and fully as possible. PECE also allows researchers to experiment with new
718 ways of publishing their work to address diverse audiences, expanding
719 multimodal anthropology through the incorporation of video or audio
720 clips, images, and other media into one’s text, while allowing for field
721 notes, working memos, collaborative annotations, and similar forms of
722 “raw(er) data” to be presented as well (Fortun and Fortun 2019).

The PECE design team has also documented in writing the *design logics* 723
 guiding the configuration of the platform’s architecture; these articula- 724
 tions convey the epistemic and aesthetic assumptions we built into the 725
 digital infrastructure of the platform, assumptions that reflect our commit- 726
 ments to a particular genealogy of work in cultural anthropology emerg- 727
 ing in the late 1980s, in which we situate our own work as researchers and 728
 teachers (Fortun et al. 2017). 729

All this makes PECE an instance of what Johanna Drucker and Patrik 730
 Svensson call “middleware,” their term for widely shared (digital) techni- 731
 cal tools and infrastructures that shape not only the collective experience 732
 and communications of users but their patterns of thinking and interac- 733
 tion as well. Middleware platforms, in other words, from PowerPoint to 734
 Twitter to Wordpress to PECE, are never neutral; their design and materi- 735
 al structures shape and color the inflected meanings they enable and 736
 convey. The humanities and humanistic social sciences need to focus and 737
 develop analytic attention to middleware of all sorts, Drucker and Svensson 738
 argue, both to how it functions and how it exceeds and/or fails those 739
 functions, to better understand the ways digital infrastructure indeed 740
 structures—often invisibly—its processes and products (Drucker and 741
 Svensson 2016; see also Dourish 2017; Goldberg 2015; see Franzen, 742
 this volume). 743

What Metadata Affords in PECE 744

As users upload ethnographic data into PECE as image, text, audio, or 745
 video artifacts, they are prompted to describe each file with the 15 core 746
 metadata elements defined in Dublin Core; this includes Title, Creator, 747
 Subject, Description, Publisher, Contributor, Date, Type, Format, 748
 Identifier, Source, Language, Relation, Coverage, and Rights. To us, the 749
 complex and conflicting data ideologies that informed the design of this 750
 metadata standard position it to embody more than “recommended prac- 751
 tice” in data management. 752

At the time of Dublin Core’s design in the mid- to late 1990s, the 753
 information and computer scientists working to formalize the standard 754
 sought to define a set of “core” metadata elements that could be used to 755
 universally describe content on the newly emerging World Wide Web. 756
 Well aware of the heterogeneity of “languages” different communities use 757
 to describe their data, they were worried about how metadata standards 758
 could proliferate and thus further splinter the way diverse communities 759

760 represented their data. This became a key factor in debates that emerged
761 between what the Dublin Core community referred to (in an acknowl-
762 edged oversimplification) as the “structuralists” versus the “minimalists”
763 (Weibel et al. 1997). Structuralists argued that the metadata standard
764 should incorporate more complex structures¹⁰ so that communities using
765 different words or models to define and order their data could extend the
766 standard to meet their specific needs. Minimalists, on the other hand,
767 argued that the standard should be as simple as possible—only incorporat-
768 ing concepts for which the greatest number of people could agree upon a
769 single, stable definition. Both groups wanted to prevent the proliferation
770 of standards—the former, by adding structure to make Dublin Core more
771 flexible to diverse needs, and the latter, by restricting the standard to what
772 could be agreed upon as a core set of concepts.

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

781 For us, however, it is the very imprecision around definitions in Dublin
782 Core that makes it such a robust standard for contextualizing our data.
783 The metadata standard provides enough common structure to give
784 anthropologists and other empirical humanists accessing and interpreting
785 shared data some context as to where the data came from and who should
786 be credited for its creation, contribution, and publication. However, we
787 also acknowledge that, at least when it comes to anthropological data, set-
788 ting canonical definitions for seemingly “standard” terms like “creator,”
789 “rights,” and “subject” is a deeply political act, demanding that one takes
790 a stance on what it means to create cultural narratives, what it means to
791 claim ownership of those narratives, and what it means to impose a classi-
792 fication on them. In prompting users to fill out metadata profiles accord-
793 ing to Dublin Core standards in PECE, we have sought to leverage the
794 standard’s looseness and imprecision (design affordances that emerged
795 from the conflicting data ideologies that constituted it) to trouble what it
796 means to date, geographically pinpoint, credit, describe, or assign owner-
797 ship to ethnographic data—all acts that were critiqued as part of
798 anthropology’s increasing entanglement with poststructuralist, feminist,
799 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 re-interpreted. 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.¹¹

840 METADATA AND THE SEMIOTIC INFRASTRUCTURE TO COME

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

852 The diverse, expanding, and somewhat disorienting metadata universe
853 depicted by Riley at the opening of this chapter, therefore, is almost cer-
854 tain to become more diverse and more expansive but perhaps, paradoxically,
855 less disorienting as anthropologists—methodologically committed as
856 we are to inviting excess and confusion, letting them illuminate and dis-
857 rupt established habits, and, over time, re-figuring new ones in shared
858 intersubjective conversation with others—become more adept through
859 extended involvement in these new digital infrastructures and domains.
860 Metadata models and practices that have been vital to the characterization
861 and stabilizing preservation of (a small fraction of) cultural anthropologi-
862 cal data will have to be extended, augmented, or otherwise re-formed or
863 re-placed. We’ll need to invent new metadata forms and practices better
864 suited to making our data more easily but justly shareable, more discover-
865 able in wider and more diverse and even unknown use-contexts. Our kind
866 of qualitative data and analyses—special enough without needing to be
867 extraordinarily or confoundingly so—needs to be disseminated in new
868 ways, among researchers familiar and strange, and between researchers
869 and new publics. Our kind of qualitative data and analyses need to be re-
870 interpreted in new ways, made available for re-use and re-analysis, more
871 (but not necessarily fully) open so that it can be mashed up and meshed
872 with other emergent data. It requires building new digital infrastructures,
873 like the PECE and Mukurtu¹² platforms openly available on Github, to
874 make new archival places open to more anthropologists rather than the
875 few (elite, elder) beneficiaries of such limited sanctioned institutional data
876 infrastructure as exists for conserving their data and knowledges.¹³

877 Every day, in numerous contexts, anthropologists old and young are
878 generating new data and new analysis about human situations that, like

memory in Ireland, are some mix of intellectually fascinating, politically 879
 vital, and culturally urgent, in locations from the most mundane and pos- 880
 sibly imperiled to the most elite and developed but nonetheless constantly 881
 on the move, as all cultures are. These anthropologists might be involved 882
 in projects that try to understand and address “wicked” problems of sus- 883
 tainability, ecological and public health, democratic governance, and simi- 884
 larly urgent demands that transcend all disciplinary boundaries, and their 885
 data deserves and indeed needs to be more openly shareable than it is. Our 886
 infrastructure also needs to be more open to experimental tinkering with 887
 data and metadata models that capture just a bit more interpretive data 888
 about creative interpretive practice in all its fragmented, non-holistic glory. 889
 All this requires linking infrastructural work (undervalued, underfunded, 890
 and underattended to in almost all fields, but especially so in anthropol- 891
 ogy) to the cultural work of shifting the discipline’s data ideologies. 892

We’ll therefore need to invent new forms of collaboration with other 893
 kinds of semiotic infrastructure experts: software developers, web and data 894
 scientists, and similar “technical” types. Work of this kind has to be experi- 895
 mental, not in an avant-gardist sense but in one we take from the sciences 896
 themselves: making careful perturbations to functioning systems to explore 897
 their latent capacities and affordances, guided by creative insight, observ- 898
 ing and evaluating outcomes, and re-iterating the procedure. It’s the kind 899
 of work we’re still learning how to do through our continued hands-on 900
 development and design of the PECE platform—a kind of sociotechnical 901
 work we’ve learned, anthropologically, in large part through our extended, 902
 distributed, engaged conversation with our collaborator friends in the RDA. 903

Acknowledgments The development of the PECE platform and the writing of 904
 this chapter were made possible through the support of the National Science 905
 Foundation (Award #1535888), “Environmental Health Governance in Six Cities: 906
 How Scientific Cultures, Practices and Infrastructure Shape Governance Styles.” 907

NOTES 908

1. David Ribes (2019: 524) writes that “domains refer to those fields (often 909
 scientific, but not exclusively) concerned with worldly and specific matters, 910
 for example, linguistics is the ‘domain science’ of language, biologists are 911
 the ‘domain experts’ of organic life, and so on. The logic of domains parses 912
 the world into two main categories, one is either ‘in a domain’ or one is 913
 working ‘independently’ of any domain.” Typically, according to this 914

- 915 “logic of domains,” the computer, data, and information scientists design-
 916 ing data infrastructure for domain communities are considered to be work-
 917 ing independently of any domain.
- 918 2. Within the RDA, most of any group’s work, including our own, occurs in
 919 the “spare time” of individual members in contact via electronic commu-
 920 nications, but twice a year there is a large RDA “plenary” at which groups
 921 can meet face to face. We have used these plenaries (13 to date) in part as
 922 opportunities to convene sessions with other empirical humanists inter-
 923 ested in advancing data infrastructure that is attuned to the specific chal-
 924 lenges and needs of preserving and sharing such research. We have also
 925 attended the plenaries so that we can communicate our unique commit-
 926 ments and challenges to more technical groups attempting to develop data
 927 infrastructure that can facilitate data sharing across disciplinary borders.
- 928 3. When querying a particular research community’s data ideology, we con-
 929 sider questions such as:
- 930 • What does a particular research community seek to understand, and what kinds of
 - 931 data and analysis advance such understanding?
 - 932 • How does a particular research community leverage theory and comparative
 - 933 perspective?
 - 934 • What does a particular research community seek in collaboration?
 - 935 • What does a particular research community seek to accomplish through their data
 - 936 representations, and what understandings of language, knowledge, and commu-
 937 nication underpin their efforts?
- 938 4. A reductive fetishization of sharing has structural parallels with at times
 939 naive mobilizations of “ecological” approaches, or a simplistic valorization
 940 of “connectedness” that can gloss over how connections need not always
 941 be symbiotic and just; relationships can also be predatory, abusive, extrac-
 942 tive, parasitic, and so on.
- 943 5. In addition to numerous colloquia at a number of universities, PECE has
 944 been presented, reviewed, and discussed as a collaborative opportunity at
 945 meetings held by: the American Anthropological Association, the Society
 946 for Cultural Anthropology, the Society for the Social Studies of Science,
 947 the Swiss Anthropological Association, the National Science Foundation,
 948 and the Research Data Alliance.
- 949 6. In an early ethnographic moment in the prehistory of PECE, we attended
 950 an interdisciplinary conference on asthma, a complex example of “coupled
 951 human-natural systems” par excellence, and learned about just how diffi-
 952 cult this interdisciplinary “team science” can be in practice. This ethno-
 953 graphic moment sparked our sense of the need for The Asthma Files
 954 (theasthmafiles.org), the first instance of PECE before it was formalized
 955 into a stand-alone digital infrastructure that can be downloaded and
 956 installed to support a range of ethnographic projects. In this case, form
 957 (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.
7. 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.
 8. Sutcliffe-Braithwaite writes that "it can sometimes be impossible to recover all the contextual information surrounding a particular interview. Yet it is still possible to re-use 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.
 9. 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 re-analyzes here.
 10. This division primarily manifested in debates over whether users should be able to "qualify" metadata fields—that is, whether they should be able to attach additional attributes to 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.
 11. On "civic science," see Fortun and Fortun (2005).
 12. 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.
 13. 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 *Re-interpretable*, 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

1002

- 1003 Allen, A. P. (2018). *Recall Initiative: From Memory and Life History, to Ireland*
 1004 *and History*. Digital Repository of Ireland [Distributor], Irish Qualitative Data
 1005 Archive [Depositing Institution]. <https://doi.org/10.7486/DR.I.wp98p078n>
 1006 Beall, J. (2004). Dublin Core: An Obituary. *Library Hi Tech News*, 21(8), 40–41.
 1007 Borgman, C. L. (2012). The Conundrum of Sharing Research Data. *Journal of the*
 1008 *American Society for Information Science and Technology*, 63(6), 1059–1078.
 1009 Crowder, J. W. (2017). Visualizing Tensions in an Ethnographic Moment: Images
 1010 and Intersubjectivity. *Medical Anthropology*, 36(6), 584–601. [https://doi.org/](https://doi.org/10.1080/01459740.2017.1315572)
 1011 [10.1080/01459740.2017.1315572](https://doi.org/10.1080/01459740.2017.1315572).
 1012 Dourish, P. (2017). No SQL: The Shifting Materialities of Database Technology.
 1013 In *The Stuff of Bits: An Essay on the Materialities of Information*. Cambridge,
 1014 MA: MIT Press. <https://ieeexplore.ieee.org/document/7932804>
 1015 Drucker, J., & Svensson, P. (2016). The Why and How of Middleware. *Digital*
 1016 *Humanities Quarterly*, 010, 2.
 1017 Edwards, P., Mayernik, M. S., Batcheller, A., Bowker, G., & Borgman, C. (2011).
 1018 Science Friction: Data, Metadata, and Collaboration. *Social Studies of Science*,
 1019 41(5), 667–690.
 1020 Ember, M. (2000). *HRAF in the 21st Century: Blueprint for the Future After Fifty*
 1021 *Years*. New Haven: Human Relations Area Files, Inc. [https://hraf.yale.edu/](https://hraf.yale.edu/wp-content/uploads/2014/11/HRAF-in-the-21st-Century.pdf)
 1022 [wp-content/uploads/2014/11/HRAF-in-the-21st-Century.pdf](https://hraf.yale.edu/wp-content/uploads/2014/11/HRAF-in-the-21st-Century.pdf)
 1023 Fassin, D., & Rechtman, R. (2009). *The Empire of Trauma: An Inquiry into the*
 1024 *Condition of Victimhood* (trans: Gomme, R.). Princeton: Princeton
 1025 University Press.
 1026 Fortun, K., & Fortun, M. (2005). Scientific Imaginaries and Ethical Plateaus in
 1027 Contemporary U.S. Toxicology. *American Anthropologist*, 107(1), 43–54.
 1028 Fortun, M., Fortun, K., & Marcus, G. E. (2017). Computers in/and Anthropology:
 1029 The Poetics and Politics of Digitization. In L. Hjorth, H. Horst, A. Galloway,
 1030 & G. Bell (Eds.), *The Routledge Companion to Digital Ethnography* (pp. 11–20).
 1031 New York: Taylor & Francis.
 1032 Fortun, M., & Fortun, K. (2019). Anthropologies of the Sciences: Thinking
 1033 Across Strata. In J. MacClancy (Ed.), *Exotic No More: Anthropology for the*
 1034 *Contemporary World* (2nd ed.). Chicago: University of Chicago Press.
 1035 Goldberg, D. (2015). World as Platform—Genres of Scholarly Knowledge
 1036 Production—Medium. *Genres of Scholarly Knowledge Production* (Online
 1037 Collection). [https://medium.com/genres-of-scholarly-knowledge-produc-](https://medium.com/genres-of-scholarly-knowledge-production/world-asplatform-da7f8a1f042e)
 1038 [tion/world-asplatform-da7f8a1f042e](https://medium.com/genres-of-scholarly-knowledge-production/world-asplatform-da7f8a1f042e)
 1039 Greenberg, J. (2009). Metadata and Digital Information. In *Encyclopedia of*
 1040 *Library and Information Sciences, Third Edition* (pp. 3610–3623). New York:
 1041 Taylor and Francis. <https://doi.org/10.1081/E-ELIS3-120044415>.

- Helmreich, S. (2007). Induction, Deduction, Abduction, and the Logics of Race and Kinship. *American Ethnologist*, 34(2), 230–232. <https://doi.org/10.1525/ae.2007.34.2.230>.
- Keane, W. (2018). On Semiotic Ideology. *Signs and Society*, 6(1), 64–87. <https://doi.org/10.1086/695387>.
- Knorr-Cetina, K. (2009). *Epistemic Cultures: How the Sciences Make Knowledge*. Cambridge, MA: Harvard University Press.
- Leopold, R. (2013). Articulating Culturally Sensitive Knowledge Online: A Cherokee Case Study. *Museum Anthropology Review*, 7(1–2), 85–104.
- Marsh, D. E., & Punzalan, R. L. (2019). Studying and Mobilizing the Impacts of Anthropological Data in Archives. In J. Crowder, M. Fortun, R. Besara, & L. Poirier (Eds.), *Anthropological Data in the Digital Age*. London: Palgrave Macmillan.
- Nader, L. (1972). *Up the Anthropologist: Perspectives Gained from Studying Up*. <http://www.eric.ed.gov/ERICWebPortal/detail?accno=ED065375>. Accessed 12 Oct 2012.
- Nafus, D. (2017). Data Friction. *Platypus: The CASTAC Blog*. <http://blog.castac.org/2017/02/data-friction/>. Accessed 7 Mar 2019.
- O'Donnell, K., Pembroke, S., & McGettrick, C. (2015). *Magdalene Oral History Collection*. Digital Repository of Ireland [Distributor], Irish Qualitative Data Archive [Depositing Institution]. <https://doi.org/10.7486/DRI.dn39x152w>
- Ribes, D. (2019). STS, Meet Data Science, Once Again. *Science, Technology, & Human Values*, 44(3), 514–539.
- Riley, J. (2009–2010). *Seeing Standards: A Visualization of the Metadata Universe*. <http://jennriley.com/metadatamap/>. Accessed Mar 2019.
- Roy, A. (2011). Urbanisms, Worlding Practices and the Theory of Planning. *Planning Theory*, 10(1), 6–15.
- Silverstein, M. (1979). Language Structure and Linguistic Ideology. In P. Clyne, W. F. Hanks, & C. L. Hofbauer (Eds.), *The Elements: A Parasession on Linguistic Units and Levels* (pp. 193–247). Chicago: Chicago Linguistic Society.
- Weibel, S., Iannella, R., & Cathro, W. (1997, June). The 4th Dublin Core Metadata Workshop Report. *D-Lib Magazine*. <http://mirror.dlib.org/dlib/june97/metadata/06weibel.html>
- Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., et al. (2016). The FAIR Guiding Principles for Scientific Data Management and Stewardship. *Scientific Data*, 3(160018), 1–9. <https://doi.org/10.1038/sdata.2016.18>.
- Woolard, K. A., & Schieffelin, B. B. (1994). Language Ideology. *Annual Review of Anthropology*, 23(1), 55–82.