Received: 17 May 2023 | Accepted: 1 November 2023

DOI: 10.1111/napa.12208

ORIGINAL ARTICLE

How to avoid the “infrastructural blues”? Studying-while-caring for data stewardship

Luis Felipe R. Murillo

Department of Anthropology, University of Notre Dame, Notre Dame, Indiana, USA

Correspondence
Luis Felipe R. Murillo, Department of Anthropology, University of Notre Dame, Anthropology, Notre Dame, IN 46556, USA.
Email: lrosadom@nd.edu

Funding information
National Science Foundation, Grant/Award Number: 2022639

Abstract

When it comes to climate crisis research, current debates are increasingly thematizing the needs but also the challenges of collaborative, transdisciplinary work. Geophysical characterizations of climate change are increasingly deemed insufficient to respond to the challenges that vulnerable communities face worldwide. In this paper, I describe the work of studying-while-caring for an environmental data infrastructure in order to address this issue. I suggest framing “data management” anthropologically as a question of collective stewardship that is better conceived as a “knowledge infrastructure” (Edwards 2010) instead of a formal approach to automated data curation. To examine the sociotechnical blindspots of data management, I elaborate on the anthropological concept of “infrastructural blues” based on the data engineering work I conducted. For the conclusion, I discuss the concept of “common” as a substitute for “open” technologies and address the broader implications of the proposed shift toward community stewardship and self-determination as guiding practices for socio-environmental data governance.

KEYWORDS
digital anthropology, data management, data stewardship, open technologies

INTRODUCTION

When it comes to climate research, current debates are increasingly thematizing the needs and the challenges of collaborative, transdisciplinary work. Geophysical characterizations of climate impact are increasingly deemed insufficient to respond to the challenges that vulnerable communities face worldwide (Crate & Nuttall, 2016; Lahsen & Turnhout, 2021; Reyes-García, 2016; Robards et al., 2018). It is in this context that new collectives are being formed with researchers, technologists, and climate-impacted community organizations to create new methodologies, datasets, analysis and visualization tools, and infrastructures for collaborative socio-environmental research (Eicken, 2021; Pulsifer, 2012).

In this article, I describe the experimental work of studying-while-caring for an environmental research infrastructure. This work was conducted as a contribution to the NSF-sponsored project "Understanding the Changing Natural-Built Landscape in an Arctic Community," led by an interdisciplinary team at University of Virginia (UVA) for the study of the impact of climate change on the built infrastructure in Utqiagvik, Alaska. The research team was composed of environmental scientists and social scientists in addition to local Indigenous companies, organizations, and government authorities (such as Tribn, UIC-Science, Taġiuġmiullu Nunamiullu Housing Authority, and the North Slope Borough Department of Planning and Community Services). I was invited to join the project as a technologist with expertise in community-based environmental sensing networks. But, as soon as I started to learn about Arctic science debates on "co-production of knowledge," I realized that I could contribute as well as an anthropologist by examining questions of data protection, openness, and reciprocity, while conducting and translating the technical work that was necessary for the project. I took on the responsibility of organizing the environmental data that was generated by a micro-meteorological sensor network. My work involved the design and implementation of a research data workflow in consultation with project members and local organizations, which I visited to present the work I report on here and inquire about the environmental issues they perceived as the most urgent. In this process, I realized that an anthropological approach was urgently needed to examine the exchange across sociotechnical differences of knowledge and expertise, interpreting relations and tensions in data management in the register of the gift (Caillé, 2007; Mauss, 1985).

Critical voices are joining in unison to demand from Arctic scientists that listening and reciprocating through collaborative research is more important than collecting data and publishing on impacted environments, which has historically characterized the "normal" technoscientific attitude (Erickson, 2020; Smith, 2012; Yua, 2022). "Collaboration," "participation," and "co-production" are disputed keywords, raising, at once, hopes and red flags, triggering historical wounds and inter-generational trauma for Arctic communities in their interactions with Southern colonial enterprises (Stevenson, 2012; Zanotti et al., 2020). In this context, the National Science Foundation "Navigating the New Arctic" (NSF-NNA) program is far from neutral. After I accepted to join the project with Arctic scientists, the first document I read was a protest letter drafted by Indigenous organizations (Kawerak inc., Association of Village Council Presidents, Bering Sea Elders Group, and Aleut Community of St. Paul Tribal Government). The signatories write: "We appreciate the recognition in the NNA request for proposals (RFPs) of the need for research in a rapidly changing Arctic to take a co-production of knowledge (CPK) approach. However, there was no meaningful effort at CPK as far as we have seen for these proposals" (Kawerak, 2020, 7). At an elementary anthropological level, the message delivered was loud and clear: scientists have *taken but have not given back*. History repeating itself, except this time with added layers of complexity and abuse concerning the production, use, distribution, control, and attribution of data.

To redress historical wounds and engage in reparation, the NSF funded the "Navigating the New Arctic Community Office" (NNA-CO) as a hub to support collaborative research. To help realize what researchers saw as a promise of "co-production of knowledge," the NNA-CO has been proactive in bringing community members, scientists, artists, and students together to create ties that may serve to interrupt extractive practices of the technosciences. While it is still early to evaluate the outcome of the NNA-CO, its community events have created space for powerful listening sessions, increasing awareness of the importance of ties of respect and trust for meaningfully "co-productive" research. In the very first NNA-CO meeting in 2020, the Iñupiat artist Joseph Senungetuk described to the scientists in the audience the diary of his father, a subsistence hunter who kept a detailed record of his activities, starting with a daily account of environmental conditions. On the slide, the participants of the NNA-CO meeting could see one of Senungetuk's paintings with the representation of a Iñupiat "story knife" (a story-telling device) above the sentence: "Not high tech. The story knife sufficed for centuries to help us remember the important ways of knowing. Colonization came like a giant eraser." In the context of the "community office," the call helped scientists to understand that forms of knowledge exchange through story-telling were key for self-sufficient Arctic ways of life. During Q&A, another Iñupiat resident of the North Slope of Alaska added that the presumption of "newness" of the NSF program was itself grounded in a colonialist attitude and trope of the "new frontier;" while another community member reminded the environmental scientists that data collection from native environments should be understood as Indigenous data, and, therefore, subject to Indigenous rights to self-determination and tribal government oversight.

In what follows, I will proceed by describing the steps I took to avoid falling into the trap of data extraction as an anthropologist-turned-technologist. I will explain how I reframed data management as a problem of community data stewardship that is better examined as a "knowledge infrastructure"—here understood as a "network of people, artifacts, and institutions that generate, share, and maintain specific knowledge about the human and natural worlds" (Edwards, 2010, 17). To examine the challenges and potentialities of this alternative framing, I discuss the notion of "infrastructural blues" that stems from the experience with unequal power dynamics in scientific data management. I further describe the efforts of working across conflicting demands for the integration between FAIR data guidelines and CARE data principles in the space of heightened epistemic conflict. FAIR is an acronym used by the data management community that stands for "Findable, Accessible, Interoperable, and Reproducible" data (Wilkinson, 2016). The basic goal of FAIR is to facilitate what is called "computer-actionable," automated data management, given the rapid increase in size, scale, velocity, and variability of data collection. CARE, on the other hand, stands for "Collective Ownership, Authority to control, Responsibility, and Ethics" and concerns Indigenous data governance with an emphasis on aspects of data ethics that cannot be automated. CARE has been an influential reaction to the shortcomings of FAIR in the context of data stewardship (Carroll et al., 2020), hence its central importance for the promotion of community data stewardship more broadly. To engage this debate, I will foreground the anthropological aspects of data engineering through which technologists feel (and empathize with those who feel) the "blues" of exploitative, data-extractive practices. This is meant to be a tactic positioning, I suggest, for situating us in the context of everyday data practices where "data standards" are not so standard; where "data sharing" guidelines are imposed and often disputed, where "openness" in sociotechnical arrangements is hard won, when it is achieved at all; where the "blues," in sum, strike those who are made invisible behind data stores, pipelines, and management plans. The technopolitical invisibilities and impossibilities created by the "giant erasers" of large-scale data collection and processing can be challenged, I argue in conclusion, at various aspects of data work to promote collaborative futures for socio-environmental research.

FROM DATA MANAGEMENT TO DATA STEWARDSHIP

Data management (DM) is a curious domain of sociotechnical work on information infrastructures, architectures, standards, and interfaces often identified, after Susan Leigh-Star (1999), with the study of "boring things." DM concerns the application of expert knowledge to the machine-actionable administration of technical objects: from forms of representation (as "datum") to forms of indexical representation about representations ("meta-datum") that are encompassed by human-plus-machine "management policies" for data disposition, integrity, access, protection, preservation, and sharing (Moore et al., 2015; Miksa et al., 2019). Data managers assume the role of those who specify, plan, regulate, prescribe, and proscribe through exercises of expert authority that does not go unchallenged, un- or misrecognized, or ignored out of sheer obliviousness in "userland" (sphere where so-called "computer users" operate) about what happens in the digital "backend"—domain of the *cryptotechnique* (of data infrastructures) that is opposed to the *phanerotechnique* of data management plans (Simondon, 2008). Respectively, the domain of the profound *technicity* to the initiated versus the one that is visible to the so-called "users." DM experts often come from engineering, computer and library science, and exchange among themselves in regular gatherings, propose and update protocols, standards, data and metadata services. They promote institutional campaigns to instruct data producers and users on how to manage themselves and their data, often with mixed results and truck-loads of frustrations. Whereas expert management is a well-established and studied modernist obsession for governing populations (rendered as data) through bureaucratic institutions, self-management has become the main orientation with populations now rendered atomized individuals, where behavioral modification is guaranteed through digital mass surveillance (Deleuze, 1992; Weber, 1922 [1978]; Zuboff, 2019). It is in this emergent global network of networks, where data represents the "machine oil" of digital surveillance platforms, that the call for data sovereignty proves to be urgent to address the pernicious effects of "Big Data."

If it is knowingly the case that management is a fundamental part of the modernist project with an underlying history of violent (data) extraction, data engineers are not left with many alternatives but to resituate and recreate their practices in the context of exchange on community data stewardship with "care" for and alongside those who are "targeted" by knowledge infrastructures (Carroll et al., 2020; Kukutai & Taylor, 2016; Lovett et al., 2019). It is in this emergent space of dispute for the digital *otherwise* that community organizers, researchers, and technologists can find the most generative reorientation of their work with digital infrastructures, mediating sociotechnical relations they entertain with their environments, themselves, and other-than-humans (Amrute & Murillo, 2020; Murillo & Dosemagen, 2021). This is the also the generative space where the question of the "digital common" can be recast under a new light: one that makes tangible the limitations of the liberal, Euro-American romance of "Free Culture"; while, at once, maintaining its critical mass for disputing intellectual property maximalism and fostering data protection and privacy *by default* for racialized communities that carry the brunt of the environmental crises (Brown, 2004; Dosemagen et al., 2021; Hayden, 2010; Nixon, 2011). "Caring for data" in this context is encompassed by other forms of care for humans and non-humans.

To illustrate this point, we can turn to a recent but transformative event that took place in the context of the Research Data Alliance (RDA) "Indigenous Data Sovereignty Interest Group" with the release of the "CARE manifesto" by a group of researchers, technologists, and data management experts (Caroll et al. 2020). CARE principles were first articulated as a response to the FAIR guidelines for "Findable, Accessible, Interoperable, and Reproducible" research data (Wilkinson, 2016). While the latter was primarily meant to be machine-actionable, the former represented first and foremost, a technopolitical demand for entering and cultivating careful relationships that are not amendable to digital automation. If FAIR was meant to facilitate the automation of very complex data management tasks, CARE represented a call for establishing Indigenous communities' control over data governance matters. Recent efforts have been directed by the original CARE proponents to, not only operationalize CARE with FAIR guidelines, but to advance other technical aspects of data management, such as the specification of a provenance metadata standard (IEEE "Recommended Practice for Provenance of Indigenous Peoples' Data") that requires the contextualization of data from Indigenous and other marginalized groups (Caroll et al., 2021). CARE offered, overall, an integrative proposal: "Given the tension between protecting Indigenous rights and interests in data while encouraging FAIR data in a global research environment that also supports open data," the CARE proponents write, "implementation of the CARE Principles should be seen as a required dimension of open and FAIR data that ensures the use of data aligns with Indigenous rights, is as open as determined by Indigenous communities, is purposeful, and enhances the wellbeing of Indigenous Peoples" (Carroll et al., 2022, 2). It is crucial to emphasize here that, while integrative efforts can be extended to non-Indigenous colleagues working on data modeling and engineering problems, the work of CARE for Indigenous data must be performed by Indigenous engineers and scientists for and by themselves as Carroll et al. (2022) remind us.

Another key point concerns CARE in the public debate on data governance: while its pressing need is indisputable, its realization is far from straight-forward. One of the major difficulties lies in the details of the actual work of integration with FAIR guidelines in a terrain of dispute over what counts as carefully contextualized data practices, bringing about the "infrastructural blues" we feel when we take into consideration the disconnect between opposing moral economies of technopolitical projects, the utilitarian and the relational, the legal frameworks for individual versus collective rights, but also, the unprecedented level of concentration of power among data experts and digital platform owners in contrast with Indigenous and other racialized communities. A peculiar kind of "blue," I would add, is experienced by those who operate in this problem space with a paradox at hand. The more data management is advanced, at the level of the actual data practice, the more technologists or scientists tend to alienate those who need the most to be part of governance decisions.[1] The question, then, becomes: how can technoscientific experts embrace collective data governance practices? This is the point where anthropological sensibilities for *studying-while-caring* for data stewardship can be of great help.

DATA INFRASTRUCTURE AS AN ANTHROPOLOGICAL PROBLEM

In 1974 Roberto Da Matta published his essay "The Ethnographers' Craft, or How to have Anthropological Blues" which soon became one of the most important introductory texts in Brazilian anthropology. In his piece, Da Matta approaches the ethnographic experience through ritual analysis, describing it in three phases: (1) preliminary, or "intellectual phase," where anthropologists learn from publications on theory and method; (2) intermediary or "practical phase" where researchers are fundamentally occupied by everyday matters (such as obtaining visas, permits, figuring out where to stay, who to contact, etc.); and, finally (3) personal or "integrative phase," in which "we must," in Da Matta's words, "synthesize biography with theory, and the practice of the world with that of the (ethnographic) craft." These phases are well-known to any practicing ethnographer, which are punctuated in the personal or integrative phase by something that Da Matta calls "anthropological blues." The "blues" was first suggested to the author by Jean Lave, based on her experience in the Amazon that involved the most personal aspects of ethnography, such as the overpowering sentiment we feel as we enter intricate meshes of relations, but also of improvisation that leads to the most important insights about what it means to be human alongside other humans and other-than-humans. The "anthropological blues," Da Matta observes, "insinuates itself in ethnographic practice unexpectedly" (op. cit., p. 30, *my translation*). The "blues" opens up the experience of participant observation to human struggles and affects, solidarities and disaffects.

In data management, a similar experience seizes upon the technologist with the overpowering sentiment of helplessness due to inter-incomprehension and, often, isolation as one attempts to take possibilities for collaborative work seriously. The data engineer treads through similar stages of: (1) ideation (with "over-abstraction" of data problems); (2) realization of practical urgencies; and (3) experimentation of the blues of data protection and sharing (that may never happen). To transpose the historical metaphor to another context of encounter across power differentials, the "infrastructural blues" can be analogized with the descent of the folkloric "blue devils" at the level of the digital interface that some of us feel while working, for example, on a data management plan that is supposed to be executed as it was written (with its human and machine-actionable provisions); while a very few outside information science circles can actually understand what the effort actually entails. What exclusions, occlusions, and erasures are (re)produced through digital media in the process? From the existing literature on data infrastructures, we learn that large-scale scientific collaborations have been identified by ethnographers as sites of sociotechnical breakdowns and standard fictions (Bowker et al., 2010; Edwards et al., 2011; Ribes & Jackson, 2013; Star, 1999). It is well-documented in the infrastructure studies literature, in particular, the phenomenon Paul Edwards (2010) called "data frictions" of climate modeling which accounts for the multiple determinations (social, technical, institutional) of data as it moves about from instruments to databases, from computer consoles to published papers. The work of data engineering that is required involves communicating through standards, specifications, management plans, source code, repositories, data types and formats, and interfaces for making applications "talk to each other" through Application Programming Interfaces (APIs). It demands domain knowledge that is perceived to be sufficient to operationalize a scientific "workflow" into a running "data pipeline," but, for the purposes of community data stewardship, it hardly is.

Paradoxically, the better a data engineer aligns oneself with the technical requirements of data management, the stronger is the tendency to alienate those who need to be part of data governance since the most important condition is absent: the distribution of power (with decentralized technical expertise and infrastructural capacity). This is always much easier said in data management plans than done, and, evidently far from good enough when it comes to the goal of supporting data sovereignty projects. FAIR without CARE can lead to extractive practices. Data openness without data literacy can be moot, because the *common*—as a political principle of communality—does not exist to mutually benefit those involved (Dardot & Laval, 2014; Federici, 2018). Despite these hard challenges, there are practical steps that data engineers can take to support community-driven data stewardship. In what follows, I will describe these steps through the lessons I learned the hard way to the tune of the "infrastructural blues."

STEPS TOWARD A COMMON ENVIRONMENTAL DATA STACK

When searching for open technologies that could be adopted for integrating FAIR principles in very-small-to-very-large data infrastructure projects I found the HDF Group, a team of data experts that have been practicing "Open Science" for 2 decades *avant la lettre*. Historically, the HDF Group has been very important for the environmental sciences for creating a file format for large-scale data storage with rich metadata ("Hierarchical Data Format," HDF) but also, more recently, a data service ("Highly Scalable Data Service," HSDS) for scientific computing. The HDF Group has its roots in the National Center for Supercomputing Applications at University of Illinois (NCSA), known for its contributions to the further development of CERN web technologies, such as the web browser NCSA Mosaic. The history of the HDF format specification is connected with the history of Mosaic since both came from the same group. Mosaic was the first browser to display images in hypertext, and HDF was first envisioned as a data sharing and visualization format. HDF derived from the "All-Encompassing-Hierarchical-Object-Oriented" format (AEHOO), a humorous acronym that, as Mike Folk identified through his efforts of digital archaeology,[2] appears in the meeting minutes of the earliest specification engineers. As early as 1987, the group had already described the need for a new standard format that would have "extensible data storage mechanism, speed, options for compression, and clear capability for gridded data storage" (Folk 2010). In the following meeting, AEHOO was proposed with key features that would be later developed with the HDF format, such as platform portability, focus on research applications, and ability to be extended to include several data formats within one AEHOO file. Here we can see what HDF would soon become in response to the requirement of science funders for large-scale data preservation. By 1988, the HDF specification (version 1.0) was first drafted to provide a standard container for data sharing. A series of HDF-based tools for visualization and (remote, collaborative) manipulation of data tables was developed and released as NCSA software.

From its early history to the present, the HDF specification has gone through several versions, extensions, and improvements to accommodate the needs of large-scale scientific data management. In its current implementation, HDF version 5 introduced the "user block," the first section of an HDF file meant for storing metadata. At a micro-sociological level, this extension was crucial for rendering (partially) visible the labor that goes into preparing a data store. This is also the space for encoding key information about context as "data provenance"—since HDF files are meant to circulate, but not without proper care in observance of data governance guidelines from their communities of origin. This is a small but important step as "provenance" encoding is fundamental for preventing communities from having their data taken, used, and circulated without their consent. We need to know how the data was produced, where it came from, what changes it has gone through over time, and who should we contact if we have questions. This rather arcane technical feature is crucial when it comes to one of the principles articulated in the CARE manifesto, "Authority to control."

For the purposes of environmental monitoring, it is common to find digital and analog instruments that have their data aggregated in "comma-separated values" (CSV) files. While this format may be sufficient for local experiments by small teams, it is often not suitable for data sharing, archival, or preservation as no contextual information is included. When engaging discussions about the comparison between HDF and other formats for data storage, I presented my interlocutors with a question: for someone who is not part of this data management group, what additional information would be necessary to help make sense of an observation?

In interpretative data modeling work, we find "(technical) specifications all the way" down in nested structures of abstraction (with their corresponding material and computational support). The digital representation of a single field "observation" on Figure 1 for a "water level" sensor is encoded in a serialized format (JSON) with a "dictionary" structure, where both data and metadata are presented in key-value pairs. Data comes with a metric in the "standard system of units" ("si_value"), but also converted to the US unit (for the convenience of users in the United States who have no scientific application for the data). The original "observation" is presented with rather limited metadata: a "datetime" string with the timestamp of data acquisition, persistent identifiers for the monitoring station ("logger_sn") and its attached sensor serial number ("sensor_sn") as well as other identifiers that pertain to a (non-standardized) ontology defined by the manufacturer.

By going through every field in the example above, it becomes clear that elementary contextual information is missing (let alone any relevant ethnographic data). Some of the missing information is represented inside the bracket of Figure 1: where is the sensor located? How

```
"observation_list":
  {
    "logger_sn": "99603325",
    "sensor_sn": "99508399-4",
    "timestamp": "2019-11-20 00:00:00Z",
    "data_type_id": "1",
    "si_value": 0.9995219339475939,
    "si_unit": "meters",
    "us_value": 3.2792714368359346,
    "us_unit": "feet",
    "scaled_value": 0,
    "scaled_unit": null,
    "sensor_key": 9658456,
    "sensor_measurement_type": "Water Level"
  }
```


METADATA:
Location?
Placement?
Sensor specs?
CF standard_name?
min_value?
max_value?
...

FIGURE 1 What is missing from individual environmental observations?

is it placed with respect to the source under observation? What are the sensor characteristics (i.e., operational resolution, precision, accuracy, drift, thermal profile, cross-sensitivities)? If the measurement has been standardized by the climate science and informatics community, using, for example the “Climate and Forecast Metadata” standard (CF), what is the “standard_name” for the measurement? What is the standardized unit of measurement? If standards have not been yet established, how do we communicate our research needs to the standardization community? Questions such as these demand and expect a particular kind of expertise that is often found at the intersections of informatics and environmental sciences. It is with this set of questions that I started the process of data management and documentation with the Arctic research group. It was in the context of this experience that I realized the need for translation and explanation of technical details, so other aspects of FAIR and CARE integration could be placed in the environmental scientists’ agenda.

In data management circles, “rich metadata” is a popular boundary-object that describes conditions for data preservation and sharing (Star, 2010). It is also one of the most emphasized aspects of FAIR. Yet, the work of metadata description is often perceived as time consuming, when not ignored outright as a waste of time. “Countless discussions of the importance of metadata in the literature reflect the nearly insuperable difficulty of getting research scientists to record even the most basic metadata,” report a group of expert infrastructure ethnographers, “let alone the meticulously detailed descriptions needed for long-term, multidisciplinary data sharing. For example, some of our interviewees estimated that up to two full days of work is required to fill in the metadata questionnaire for each of the hundreds of model runs (simulation datasets)” (Edwards et al., 2011, 673). At a most elementary level of scientific labor, sensor maintenance is hard as well. “Scratch the silicon surface,” writes another group of expert infrastructure researchers, “and you will uncover a frustrated field technician recalibrating a vandalized weather monitoring station for the third time that month” (Ribes & Jackson, 2011, 152). These are well-known sources of “infrastructural blues” for “invisible technicians” (Shapin, 1989). They signal that a very important group has been left out of the technoscientific project. As described by the CARE principle of “Collective benefit,” not only scientific advancement must be considered, but, first and foremost, “inclusive development,” “citizen engagement,” and, finally, “equitable outcomes” (Carroll et al., 2020).

Once the need for collective support for the hard work of data curation, protection, and sharing is realized, hope for integrating CARE and FAIR guidelines can be regained with the inclusion of project participants that may not be data experts themselves. Some practical steps can be taken in this regard by expanding “contextual information” of a dataset, not only information that only technoscientific experts can interpret. Another step concerns the recognition of the labor that goes into preparing a dataset by translating the work with “data storytelling” sessions with the concerned public (Gabrys et al., 2016). When combined, these

```
schema: hobo-sensor
sn: 21187245-1
sn_logger: 21401800
schema_version: 1.0
active: True
long_name: "Air Temperature"
standard_name: air_temperature
sensor_type: "12-bit temperature sensor"
model: HOBO RXW-THX-900
lat: 71.292183
lat_units: degrees_north
lon: -156.778717
lon_units: degrees_east
height: 2.0828
height_units: m
angle: 0.0
angle_units: degrees
altitude: 0.0
altitude_units: m
battery_level_max: 4.2
battery_level_min: 3.0
battery_units: v
accuracy: "±0.25°C from -40° to 0°C"
resolution: "0.02°C"
drift: "<0.01°C"
datasheet: https://www.onsetcomp.com/files/manual_pdfs/22242-F%20RXW-THC%20Manual.pdf
measurement_type: Temperature
measurement_min: -40.0
measurement_max: 75.0
measurement_units: C
missing_values: -999.9
~
```


FIGURE 2 Plain text file with sensor metadata.

steps can help to ward off the "anthropological blues," while helping to unblock a fundamental space of debate and deliberation about the goals and parameters of an environmental study. In the particular case of the Arctic research project, as we discussed levels of precision, accuracy, granularity, and specificity the environmental scientists required, I took the opportunity to also pose fundamental questions concerning what should be studied as a priority, where, for whom, with whom, and for what purposes? It is well-known in Arctic science that local community members have been historically excluded from the process of project design and agenda setting, and employed as support personnel whose knowledge was identified as "anecdotal" and irrelevant for studying climate change (Erickson, 2020; Zanotti et al., 2020). This is yet another source of "infrastructural blues" that stems not only from the experience of inter-incomprehension among technoscientific experts, but also from the experience of exclusion for those on the other side of the symbolic divide between scientists and non-scientists.

As a contribution to the Arctic research group, John Readey (from the HDF Group) and I wrote an application called "hobo-request" to facilitate the task of obtaining sensor data, organizing it, and attaching metadata by filling out plain text files collaboratively (Figure 2). My primary motivation was to have an interface that could be used in workshops to engage the collective exercise of metadata description. And so, I proceeded to discuss with research members and collaborators in events, meetings, and conferences what pieces of contextual information had to be included in our data store. In retrospect, this was a small but critical contribution toward rendering visible the invisible aspects of our work.

The HDF5 data format is composed of three parts that accommodate heterogeneous data structures in the same file: (1) "databases" as homogeneous arrays of data; (2) "groups" for storing and organizing heterogeneous databases; and (3) "attributes" to annotate databases and groups, as well as the "root" ("user block") of the HDF5 file itself (Collette, 2013). To generate metadata attributes, we read instrument datasheets and engaged in discussion with research co-participants about the fields that they considered important to be added. In addition to standard metadata fields, such persistent identifiers ("sn" for "serial number") and measurement names (as "standard_name" based on the "Climate and Forecast (CF) Metadata Conventions"), we added provenance fields for identifying loggers by name and address in both English and Iñupiaq languages. In addition to the instrument operational characteristics, we included information about sensor installation for the purposes of comparison across research sites. When the "schema" (the structure of data and metadata fields and their relationships) was finally settled, we marked it with a version number so future modifications could be tracked.

FAIR guidelines were observed at every step of the data curation. Findability (the "F" in FAIR) was improved with the inclusion of metadata fields that could be harvested at the level of an institutional repository, such as the NSF Arctic Data Center while we started, most importantly, a debate about the need for data infrastructures that should be located in the Arctic for the Arctic community. We made a conscious effort to ensure that our data service could not only be accessed, but also entirely replicated on infrastructures of all sizes (from the local infrastructure of a community center to the university high-performance computing cluster). "Accessibility" ("A" in FAIR) was improved in our case through the application of a common format that is highly portable, so the data could be shared in its unprocessed format (HDF5), but also as a NetCDF4 file, which includes a standard schema that is familiar to environmental scientists. Interoperability ("I" in FAIR) was ensured with the multi-platform affordances of HDF5, so our research data could be read, manipulated, and changed for different purposes by other community organizations and research groups using open tools without proprietary, corporate lock-ins. Finally, reproducibility ("R" in FAIR) was observed by documenting all the steps of the data acquisition and processing in addition to the base technologies and their settings.

While this is all part of the data engineering work, my attention was primarily directed toward questions of collective data stewardship (which, in turn, depended on the details of our "data pipeline" and its supporting infrastructures). As soon as the data store was well-organized and annotated with its metadata, I moved on to the other aspects of the data work that consisted in providing interfaces for data access, analysis, and visualization. At this level, I was assisted again by John Readey (from the HDF Group) who implemented the Highly Scalable Data Service (HSDS) and provided us with local and remote access to our data store. The approach we took had many benefits for collaborating outside university walls as our data service was conceived to run on fairly limited infrastructures (of one or more computers) as well as on institutional computing clusters. John and I configured HSDS at University of Virginia and University of Notre Dame, but the underlying motivation was to create conditions for the same data service to be installed in Alaska in the near future, so the benefits of the large-scale data service could be governed and used to benefit local organizations, students, and governmental officers more.

The data workflow (Figure 3) was conceived and implemented with four components that can be interpreted as constitutive parts of a data governance plan: (1) data access, acquisition, and transmission; (2) storage and preparation; (3) analysis and visualization; and, finally, (4) data sharing and long-term deposit. All these components were meant to be governed by a stewardship committee. The first component requires making collective decisions about the scope of the research according to pressing environmental challenges and community needs. The second describes storing, annotating, and processing data, so they can be put to common use through analysis and visualization. For this component, our research

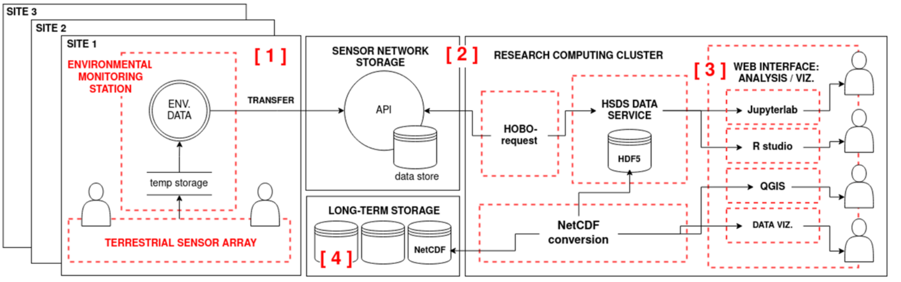

FIGURE 3 Data workflow.

project took unprocessed environmental data and extended them with metadata before they were made available for analysis. As an additional step, we are currently finalizing a NetCDF4 file converter (a popular format among environmental scientists) to process instrument data and making it available for project partners and other scientific groups (in compliance with NSF's "Open Access" mandate). Once data processing is completed, the dataset is be made available for analysis using web-based research notebooks that can be collectively created, modified, shared, and redistributed according to the community orientation toward data mutualization. Finally, in the fourth component, NetCDF files with "good enough" metadata are sent to a long-term repository.

At this high level of model abstraction, it is important to note, the data workflow looks suspiciously simple and "seamless." All the relevant sociotechnical details and frictions are hidden, just like the blue devils in minute devilish details. The struggle for getting metadata in place is often the object of what the anthropologist of science, Sharon Traweek, has called "corridor talk," where questions of community-driven data stewardship may recede from view as non-experts drown in technical detail. "Openness" does not look as difficult to achieve when it is perceived to be a simple "technical choice." It may even appear as a "given" since everything that *infrastructures* our infrastructure (at the level of digital technologies) is mostly based on Free and Open Source technologies that can be "automagically" used. In the context of the Arctic project, however, the "blues" appeared at various stages of this process. "Openness" opened some doors for creating new collaborations, but it also closed many doors with established scientists who did not see the "incentives" to engage in exchange circuits with non-scientists. The data stewardship layer has proven to be the most difficult to bring together for the lack of interest, time, and capacity, but also due to opposition to "Open Access" in general. Among project participants, including scientists and non-scientists, there were widely divergent positions regarding the importance, the need, or the adequacy of "open data."

TWO REGIMES OF DATA GOVERNANCE: "OPEN" AND "COMMON"

The question of "openness," albeit controversial, is tightly coupled with the question of data stewardship, given the unprecedented concentration of computational power (as an extension of political power) in control of a small number of Internet companies (Srnicek, 2017; Terranova, 2022; Zuboff, 2019). "Openness" can be a deceptive notion as it can work against its intended technopolitical uses. One of the most problematic aspects of advocacy for "openness" I experienced has to do with its identification as a panacea for problems of "transparency" with complete disregard for historic inequities in terms of access to information, knowledge, and computational infrastructures. For the case I am reporting here, "openness" must be understood, conversely, as one of the key characteristics of reciprocal ties we must cultivate to render any socio-environmental project meaningfully collaborative. What is critical here is the point where the question of "open data" ethics touches upon questions of reciprocity and care for the other. According to Gilbert Simondon (2008), "open technical objects" can be distinguished by one fundamental but often overlooked characteristic: "producer" and "user" must be found in the same person. This is important for our discussion because it brings up the question of alienation through data engineering. That is, it helps us to identify and cast the problem of exclusion of the "non-technical Others" (which encompasses those who do not get to learn about the digital tools and infrastructures they depend upon). At this level of ethical questioning, "open technical objects" have relational properties instantiated through gift-economies: they embody and convey, to paraphrase Simondon (op. cit.), the "respect for the other" for they can be studied, modified, shared, and extended to integrate other sociotechnical systems.

"Openness" is never an obvious point of departure, never a "given" as often suggested in the institutional (or, more recently, corporate) discourse on "Open Science." In this experiment, with studying-while-caring for an environmental data stack, it became quite clear how much it all depended upon collective cycles of gift-giving that, as the newest social movements suggest, runs on the slow temporality of relationships, not the accelerated time of commercial software development. At this sociotechnical level, the "infrastructural blues" strikes when possibilities of mutual-aid are foreclosed, when collective data work is deemed "superfluous or inefficient," when data infrastructures only serve the machine-automated use of human beings for the purposes of data extraction (Zuboff, 2019). It is also at this level that identifying, contributing, and supporting common technologies could be considered part of data stewardship. Below I will describe the blueprint for a data politics that diverts the tension involving "open" provisions for data sharing. This approach requires substituting the term "open" to qualify technical objects and practices, given the confusion it creates with utilitarian and legalistic understandings of Open Source development. This confusion introduces fundamentally the misidentification of a substantial moral economy for a software development methodology or a set of alternative open licenses for digital objects (see Coleman, 2012; Kelty, 2008; and Leach, Nafus and Krieger, 2009 for the anthropological elaboration of this point).

The problem of "openness" and access control has been navigated before by the same group of data experts who drafted the CARE manifesto, except with the difference that they have not examined it from an anthropological perspective. In one of their latest publications on CARE, they presented us with the problem of operationalization of their principles (Carroll et al., 2021). The proponents of CARE for data governance argue that their principles are meant to be complementary to FAIR principles. Yet, fundamental differences in scope of the role of data stewardship are at stake: for the former, a much broader understanding of data management is encompassed by data sovereignty that necessarily involves Indigenous communities. For the latter, the goal is to solve hard problems of scale in data management through the promotion of automated, computer-actionable data management policies. FAIR has clear limitations that have been spelled out by CARE proponents: "The FAIR Principles are aligned to the global shift towards open science and open data, promoting data centric criteria that facilitate increased data sharing among entities while ignoring relationships, power differentials, and the historical conditions associated with the collection of data." (Carroll et al., 2021, 3).

To help facilitate the work of data stewardship, I describe below a provisory schema that may serve for determining accessibility, control, and usability along three dimensions, instead of two of the usual binary between "open" and "closed / proprietary" data. It may be fruitful in the debate between FAIR and CARE to distinguish, I suggest, between open, common, and personal data (Table 1).

This basic schema suggests a "common" data access policy that encodes principles of reciprocity with community-led governance in its process of production, organization, registry, and distribution. It is selectively "open" when it needs to be for those who are responsible

TABLE 1 Governance data types.

Data type	Access policy
Open	Public domain or openly licensed data
Common	Defined by collective governance mechanisms that are community-driven and context-dependent
Personal	Pertains to persons (human or non-human) as conceived by a particular group, that is, according to the cultural and/or juridical definitions of what counts as a person; this type of data assumes privacy by default and requires meaningful consent or collective deliberation by the data stewards to be made either open or common

for its stewardship. But it is also delimited to an assigned group, not having the permission to circulate without restrictions. "Common" data mediates as a new category between "open" and "personal," helping data stewardship collectives decide what can be rendered "open" (as in public) or what needs to be kept for the persons (under strict and restricted access). The goal of using this schema is to initiate discussion and deliberation on more complex data access policies that are context-dependent and sensitive. The "common" concerns, after all, the governance of data at the local level. One example of common (not open) data can be found in the definition of "Indigenous data" provided by Carroll et al. (2021): "Indigenous data are data, information, and knowledge, in any format, that impact Indigenous Peoples, nations, and communities at the collective and individual levels; data about their resources and environments, data about them as individuals, and data about them as collectives" (1).

CONCLUSION: "STUDYING-WHILE-CARING" FOR COMMON RESEARCH TOOLS

How we can convert the "infrastructural blues" of data management into the experience of collaborative design for new ecological ways of life with socio-environmental projects that eliminate the traps of extractive technoscientific relations? Research design, like any other form of design, we are reminded by Arturo Escobar (2018), is a form of "ontological design" for particular forms of life. How can we recast in socio-environmental research, therefore, data governance as ours and not "someone's problem," a common problem for the future of common data stewardship?

I started this article with a brief reflection on the role that exchange practices have in the context of Arctic communities to devise a type of common data infrastructure and data access policy that partakes in the anthropological register of gift-giving. Albeit in a provisional form, I hope to have demonstrated that important lessons for the future of community data stewardship can be found at the basic level of the technical work with open technologies. But there is much more. Key lessons can also be found in the ongoing work of Indigenous technologists and scholars to create guidelines for common data stewardship. This is not meant to suggest the expropriation of Indigenous initiatives for non-Indigenous purposes, but to learn from them to problematize the indiscriminate practices of data extraction of Indigenous and non-Indigenous peoples. My inspiration for working in this space comes from Indigenous debates on data sovereignty, but also from Free and Open Source technology projects from the Global South where technical objects figure as gifts that are given, that can be accepted, that require engagement to be understood, modified, and shared by communities outside the main (exclusionary) axes of technoscientific development (Silva et al., 2022). To receive a gift is to enter relationships of mutualization or "commoning" (with a history of relationships that are reflected in the demands that are made upon us) and assume a form of responsibility that involves observing provisions, at the infrastructural level of any data store, for FAIR and CARE as I have demonstrated in an existing environmental data workflow.

FAIR guidelines—as exemplified in the way we care for our data (designing, documenting, implementing, depositing, protecting, and sharing)—may meet CARE principles through an economy of the gift. "Collective benefit" may be achieved through a moral economy of common data. Caring for "common data infrastructures" means in practice caring for the collective and its environment from which the data has been generated. "Authority to control" confronts us with another dimension of gift-giving: from the "open by default" of the Open Data movement to an urgent and renewed understanding of the "common" (as a political principle and operator of relationships) that concerns sociotechnical ties and not solely flexible licensing or technical affordances for manipulating "open resources" with "open tools." "Responsibility" and "Ethics," finally, may mean designing common digital objects and infrastructures with responsibility for the relationships we must cultivate in order to conduct any meaningful project in a truly collaborative register.

What I presented in this article represents the initial steps I have taken to create conditions for mutualization ("commoning") and stewardship of data infrastructures. The work is ongoing and much of the data governance aspects have yet to be established. As I discussed in previous sections, there are key aspects of data governance in debate today concerning questions of autonomy, sovereignty, and access that are still wide open. These aspects involve high-level political decisions, agreements, and treaties, but also, as I hope to have demonstrated, rely on everyday, micro-level practices of data modeling and engineering. At this level of ordinary technical practices, where "invisible engineers," such as myself (in the skin of an anthropologist), operate, it is not really up to the technoscientific experts to decide on a particular data management scheme at their universities or corporate offices far away from climate-impacted communities. Studying-while-caring for environmental data infrastructures can be one of the antidotes to the "infrastructural blues" that stem from the isolated work of data curation. There are plenty of gifts of common technology that may keep on giving if we open ourselves to acknowledge and redress the sources of the "blues" that can be heard in ongoing technoscientific projects in multiple locales and scales.

ORCID

Luis Felipe R. Murillo https://orcid.org/0000-0003-0563-7783

ENDNOTES

[1] For a telling example about this debate, see the NACER report prepared by community science activists for the Environmental Protection Agency in United States (Dosemagen et al., 2018) where questions of power and trust are discussed as key issues in the relationship that government organizations have entertained (for the purposes of environmental monitoring, but also environmental data management) with environmentally-harmed communities.

[2] Mike Folk, HDF/HDF-EOS Workshop XIV, September 28, 2010. Accessed on May 26, 2022. URL: https://vimeo.com/33973242

REFERENCES

Amrute, S., and L. F. R. Murillo. 2020. "Introduction: Computing in/from the South." *Catalyst: Feminism, Theory, Technoscience* 6(2): Available at: https://doi.org/10.28968/cftt.v6i2.34594.

Bowker, G. C., K. Baker, F. Millerand, and D. Ribes. 2010. "Toward Information Infrastructure Studies: Ways of Knowing in a Networked Environment." In *International Handbook of Internet Research*, edited by J. Hunsinger et al., 97–117. New York: Springer. https://doi.org/10.1007/978-1-4020-9789-8_5.

Brown, M. F. 2004. *Who Owns Native Culture? Cambridge*, MA: Harvard University Press.

Caillé, A. 2007. *Anthropologie du don: Le tiers paradigme*. Paris: Découverte.

Carroll, S. R., I. Garba, O. L. Figueroa-Rodríguez, J. Holbrook, R. Lovett, S. Materechera, M. Parsons, et al. 2020. "The CARE Principles for Indigenous Data Governance." *Data Science Journal*. 19:43. https://doi.org/10.5334/dsj-2020-043.

Carroll, S. R., E. Herczog, M. Hudson, K. Russell, and S. Stall 2021. "Operationalizing the CARE and FAIR Principles for Indigenous Data Futures." *Scientific Data* 8(1): 108. https://doi.org/10.1038/s41597-021-00892-0.

Carroll, S. R., I. Garba, R. Plevel, D. Small-Rodriguez, V. Y. Hiratsuka, M. Hudson, and N. A. Garrison. 2022. "Using Indigenous Standards to Implement the CARE Principles: Setting Expectations through Tribal Research Codes." *Frontiers in Genetics*. 13 https://www.frontiersin.org/articles/10.3389/fgene.2022.823309.

Coleman, E. G. 2012. *Coding Freedom. The Ethics and Aesthetics of Hacking*. Princeton, NJ: Princeton University Press.

Collette, A. 2013. *Python and HDF5*. Sebastopol, CA: O'Reilly Media.

Crate, S. A. and M. Nuttall, eds. 2016. *Anthropology and Climate Change: From Encounters to Actions*. 2nd ed. London: Routledge.

DaMatta, R. 1974. "O oficio de etnologo, ou como ter 'Anthropological Blues.'" In *Cadernos do PPGAS*. Rio de Janeiro: Museu Nacional.

Dardot, P., and C. Laval. 2014. *Commun : Essai sur la révolution du XXIe siècle*. Paris: Découverte.

Deleuze, G. 1992. "Postscript on the Societies of Control." *October* 59: 3–7.

Dosemagen, S., E. Heidel, L. F. R. Murillo, E. Velis, A. Stinson, and M. Thorne. 2021. *Open Climate Now! In: Branch. Issue 2, June 14, 2021*. https://branch.climateaction.tech/issues/issue-2/open-climate-now accessed on May 1, 2021.

Edwards, P. N., M. S. Mayernik, A. L. Batcheller, G. C. Bowker, and C. L. Borgman. 2011. "Science Friction: Data, Metadata, and Collaboration." *Social Studies of Science* 41(5): 667–90.

Edwards, P. N. 2010. *A Vast Machine: Computer Models, Climate Data, and the Politics of Global Warming*. MIT Press.

Erickson, K. S. 2020. *Successful Engagement between Inupiat and Scientists in Utqiagvik, Alaska: A Sociocultural Perspective*. MA Thesis, University of Alaska Anchorage.

Escobar, A. 2018. *Designs for the Pluriverse: Radical Interdependence, Autonomy, and the Making of Worlds*. Durham: Duke University Press.

Eicken, H., F. Danielsen, J.-M. Sam, M. Fidel, N. Johnson, M. K. Poulsen, and O. A. Lee. 2021. "Connecting Top-Down and Bottom-Up Approaches in Environmental Observing." *Bioscience* 71(5): 467–83. Available at: https://doi.org/10.1093/biosci/biab018.

Federici, S. 2018. *Re-Enchanting the World: Feminism and the Politics of the Commons*. Oakland: AK Press.

Gabrys, J., H. Pritchard, and B. Barratt. 2016. "Just Good Enough Data: Figuring Data Citizenships through Air Pollution Sensing and Data Stories." *Big Data and Society* 1–14. DOI: https://doi.org/10.1177/2053951716679677.

Hayden, C. 2010. "The Proper Copy." *Journal of Cultural Economy*, 3(1), 85–102.

Kelty, C. 2008. *Two Bits : The Cultural Significance of Free Software*. Durham: Duke University Press.

Kawerak. 2020. "*Navigating the New Arctic Program, NSF.*" March 19, 2020. https://www.aleut.com/wp-content/uploads/2022/01/NNA-Letter-With-Signatures-Final.pdf.

Kukutai, T., and J. Taylor. eds. 2016. *Indigenous Data Sovereignty: Toward an Agenda. ANU Press*. https://doi.org/10.22459/CAEPR38.11.2016.

Lahsen, M., and E. Turnhout. 2021. "How Norms, Needs, and Power in Science Obstruct Transformations Towards Sustainability." 16(2): Environmental Research Letters.

Leach, J., D. Nafus, and B. Krieger. 2009. "Freedom Imagined: Morality and Aesthetics in Open Source Software Design." *Ethnos* 74(1): 51–71. Available at: https://doi.org/10.1080/00141840902751188.

Lovett, R., V. Lee, T. Kukutai, D. Cormack, S. C. Rainie, and J. Walker. 2019. "Good Data Practices for Indigenous Data Sovereignty and Governance." In *Good Data*, edited by A. Daly, K. Devitt, and M. Monique, Amsterdam: Institute of Network Cultures Inc.

Mauss, M. 1985. *Sociologie et anthropologie*. 9e éd. Paris: PUF (Quadrige (Presses universitaires de France); 58).

Miksa, T., S. Simms, D. Mietchen, and S. Jones. 2019. "Ten Principles for Machine-actionable Data Management Plans." *PLOS Computational Biology* 15(3): e1006750. Available at: https://doi.org/10.1371/journal.pcbi.1006750.

Moore, R., R. Stotzka, C. Cacciari, and P. Benedikt. 2015. "Practical Policy." In Research Data Alliance. 10.15497/83E1B3F9-7E17-484A-A466-B3E5775121CC.

Murillo, L. F. R., and S. Dosemagen. 2022. 'Collaborative Climate Futures? Envisioning the Role of Open Data Infrastructures for Collaborative Socio-Environmental Research,' *Commonplace* [Preprint]. Available at: https://doi.org/10.21428/6ffd8432.8f2f6a1d.

Nixon, R. 2011. *Slow violence and the environmentalism of the poor*, Cambridge, Massachusetts: Harvard University Press.

Pulsifer, P., S. Gearheard, H. P. Huntington, M. A. Parsons, C. McNeave, and H. S. McCann. 2012. "The Role of Data Management in Engaging Communities in Arctic Research: Overview of the Exchange for Local Observations and Knowledge of the Arctic (ELOKA)." *Polar Geography* 35(3–4): 271–90. Available at: https://doi.org/10.1080/1088937X.2012.708364.

Reyes-García, V., A. Fernández-Llamazares, M. Guèze, A. Garcés, M. Mallo, M. Vila-Gómez, and M. Vilaseca. 2016. "Local Indicators of Climate Change: The Potential Contribution of Local Knowledge to Climate Research." *Wiley Interdisciplinary Reviews. Climate Change* 7(1): 109–24. Available at: https://doi.org/10.1002/wcc.374

Ribes, D., and S. J. Jackson. 2013. “Data Bite Man: The Work of Sustaining a Long-Term Study.” In *“Raw Data” is an Oxymoron* edited by L. Gitelman. Cambridge, MA: MIT Press: 147–66.
Robards, M. D., H. P. Huntington, M. Druckenmiller, J. Lefevre, S. K. Moses, Z. Stevenson, A. Watson, and M. Williams. 2018. “Understanding and Adapting to Observed Changes in the Alaskan Arctic: Actionable Knowledge Co-Production with Alaska Native Communities.” *Deep Sea Research Part II: Topical Studies in Oceanography* 152: 203–13. Available at: https://doi.org/10.1016/j.dsr2.2018.02.008.
Shapin, S. 1989. “The Invisible Technician.” *American Scientist* 77(6): 554–63.
Silva, T. C., L. F. Murillo, V. Tozzi, F. Caminati, A. Bonafé, J. Paixão, M. R. Arduini, D. Filho, and L. Xavier. 2022. “Computing from the South /Computação do Sul.” *Branch Magazine* (4): summer 2022.
Simondon, G. 2008. *Sur la technique*. Paris: Press Universitaire de France.
Smith, L. T. 2012. *Decolonizing methodologies: research and Indigenous peoples*. 2nd edn. New York: Zed Books Ltd.
Srnicek, N. 2017. *Platform Capitalism*. Cambridge, UK: Polity.
Star, S. L. 1999. “The Ethnography of Infrastructure.” *American Behavioral Scientist* 43(3): 377–91. Available at: https://doi.org/10.1177/00027649921955326.
Stevenson, L. 2012. “The psychic life of biopolitics: Survival, cooperation, and Inuit community.” *American Ethnologist* 39(3): pp. 592–613. Available at: https://doi.org/10.1111/j.1548-1425.2012.01383.x.
Terranova, T. 2022. *After the Internet: Digital Networks between Capital and the Common*. Cambridge, MA: MIT Press.
Weber, M. 1978/[1921-22]. *Economy and Society*, 2 volumes, G. Roth and C. Wittich (eds.). Berkeley: University of California Press.
Wilkinson, M. D., M. D. Wilkinson, M. Dumontier, I. J. Aalbersberg, G. Appleton, M. Axton, A. Baak, et al. 2016. “The FAIR Guiding Principles for Scientific Data Management and Stewardship.” *Scientific Data* 3: 160018. Available at: https://doi.org/10.1038/sdata.2016.18.
Yua, E., J. Raymond-Yakoubian, R. A. Daniel, and C. Behe. 2022. “A Framework for Co-Production of Knowledge in the Context of Arctic Research.” *Ecology and Society* 27(1): Available at: https://doi.org/10.5751/ES-12960-270134.
Zanotti, L., C. Carothers, C. A. Apok, S. Huang, J. Coleman, and C. Ambrozek. 2020. “Political Ecology and Decolonial Research: Co-Production with the Iñupiat in Utqiaġvik.” *Journal of Political Ecology* 27(1): 43–66. https://doi.org/10.2458/v27i1.23335
Zuboff, S. 2019. *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. 1st ed. New York: PublicAffairs.

How to cite this article: Murillo, Luis Felipe R. 2024. “How to Avoid the “Infrastructural Blues”? Studying-While-Caring for Data Stewardship.” *Annals of Anthropological Practice* 48: 36–51. https://doi.org/10.1111/napa.12208