UC Santa Cruz
UC Santa Cruz Previously Published Works

Title
Metadata Is Not Data About Data

Permalink
https://escholarship.org/uc/item/6j4452xb

ISBN
9781003299912

Author
Parry, Kyle

Publication Date
2023

DOI
10.4324/9781003299912-2

Peer reviewed

Metadata Is Not Data About Data | Kyle Parry, ORCID ID: 0000-0002-0919-6842

Chapter for *Decolonizing Data* (Algorithms and Society Series, Routledge, 2023)

Abstract

Several recent texts argue against the popular definition of metadata as "data about data," reframing the phrase as misleading, restrictive, and imprecise. Rather than defend the definition or determine which critique is best, this chapter argues for new approaches to the plural, ambivalent, and non-neutral concept of metadata, particularly in consideration of the rise of social media and digital surveillance, and in further consideration of enduring calls to contest and to decolonize only seemingly natural terms of debate. Adapting critical perspectives from the study of visual culture and building on Trinh T. Minh-ha's distinction between speaking about and speaking "nearby," the chapter explores conceptions of metadata beyond the frame of aboutness, introducing a method of prepositional substitution via the provocation "data near data." Whether as hashtags, Traditional Knowledge labels, or other forms, metadata is not—or is not just—data about data. Metadata is also variously expressive, reparative, distributive, and extractive.

Keywords

Do a search for the term "metadata" in the early 2020s, and you'll find that one definition continues to enjoy pride of place: "data about data." The idea is relatively simple. Metadata refers to a special type of data. What distinguishes this type of data is neither its mode (such as qualitative or quantitative) nor its status (such as raw or processed), but its role. Whether as URLs, subject headings, tags, or still other creations, instances of metadata stand in a secondary relationship to other data. They are data that describe or document things of concern, whether that means the provenance of a data set on sea levels or the timing of a readout from a distant probe or the geolocation of a customer. If there are debates in the world of metadata, then they should—according to this line of thinking anyway—center on tools and structures for managing objects and information.

As dominant as this way of looking at metadata has become, it's not the only one that's out there. In fact, what we might think of as the landscape of metadata conceptualizations is both varied and evolving. True enough that a very large subset of that landscape takes as given what I have just summarized: the idea that the now six-decade-old term "metadata" essentially refers to data that describes other data. And yet if you take the time to read across these literatures, you won't just find nuanced adjustments of that core notion, such as the adoption of variations on "information about a resource" within certain sectors of the International Organization for Standardization (Furner, 2020, p. E35). You'll also find statements of opposition. Few in number, but compelling in their reasoning, these statements contend that "data about data" is a recipe for limited thinking about a consequential phenomenon. Some even suggest that the flaws in this definition are reflective of flaws in the very idea of metadata itself: that metadata is not really the consistent and widespread thing different fields make it out to be, and that therefore we ought to use the term only sparingly, or even jettison it entirely.

In surveying these different conceptions of metadata, we can try to decide which is the most convincing and therefore should become the standard, but we can also venture into more uncertain terrain. We can be unconvinced by the options, and just as significantly, unconvinced by the idea of picking just one. The assumption would be that, with the rise of social media and widespread, automated surveillance, and in further consideration of enduring calls to contest and to decolonize only seemingly natural terms of debate, the time has come to rethink our approach to the many materials and practices that fall under the umbrella of "metadata." Taking cues and concepts from the study of visual culture, this would be a practice of critical and expansive, but not total and final, reconceptualization. Existing critiques of "data about data" would not be overcome or denied; they would be treated as resources and provocations for an underexplored line of inquiry that takes metadata's definitional uncertainty as a difficult but ultimately tractable given. The goal would not be indefinite conceptual play, nor would it be a definitive conceptual conclusion. It would be more flexible and more effective and even more just conceptions of these typically structured, typically machine-readable meta materials that append and infuse (and sometimes themselves become) what is considered primary. In short, we can ask: if metadata isn't—or isn't just—data about data, then what else is it, and why should we care?

Misleading, Restrictive, Imprecise

It's difficult to precisely place when the phrase "data about data" began to circulate, but critiques of the notion appear to be relatively recent, potentially a reflection of the increasing scope and sophistication (and skepticism) in debates on metadata. Three versions of this critique are especially noteworthy.

One appears in Jeffrey Pomerantz's widely cited *Metadata*. Early in the text, Pomerantz argues that the popular definition, while "catchy," is the "least useful" (2015, p. 19). One problem is the term "data." For Pomerantz (2015, p. 26), "data is only potential information, raw and unprocessed, prior to anyone actually being informed by it"; in other words, "data" establishes the wrong conceptual frame for metadata, because it is, in his mind, informing and being informed that really are at issue. The other problem is the word "about." Referring to subject analysis (the labor of classifying what given texts concern), Pomerantz emphasizes that "determining what something is about is subjective"; it requires not only "an understanding of that thing" but also of the "available terms" through which one can describe that thing. From Pomerantz's perspective, the phrase "about data" fails to capture this interpretive complexity.

Despite these doubts, Pomerantz ends up arguing that data about data can be "salvaged." The key, he contends, is to shift our perspective. When we are talking about things like keywords, ID numbers, and instructions for preservation, we are talking about a class of objects in the world with the distinct quality of being potentially informative with respect to other objects, of possibly telling us something we need or want to know about these other objects, whether for the purposes of finding, storing, or otherwise. From this vantage, metadata is newly conceived as "a potentially informative object" that characteristically "describes another potentially informative object." Or, in Pomerantz's subsequent and self-consciously less "clunky" version, metadata is "a statement about a potentially informative object"; for example, the presence of the name Leonardo da Vinci in a "creator" field associated with *Mona Lisa* amounts to a "potentially informative" statement that someone of this name created the painting (and, at least from this informational point of view, the painting is likewise "potentially informative").

For a second version of the critique of "data about data," the popular phrase isn't so much misleading as restrictive. For Ranger (2012), for instance, metadata is definitely "information about an object," but it is also information about how you can find that object "in a messy file structure," and it is information about how you can or ought to use that object; it is data about the use of data. For Koster (2015), the situation is more fraught. Metadata can be data about data, but this is only the case in some rather than all situations; consequently, the term is "confusing," even "superfluous." If we were to try to isolate the phenomenon, then we would have to recognize the many other things that count as metadata, from information about "virtual entities, physical objects [and] information contained in these objects" to information about "events, concepts [and] people." In other words, there is an inertia in the landscape of metadata conceptualizations, such that the memorable aphorism "data about data" serves to undermine attention to the considerable reach of the phenomenon, a reach that extends well into social, material, and historical worlds. Metadata doesn't just refer to data and objects; it refers to actions, histories, and economies—to the lives and deaths of people, places, and institutions.

The last significant line of criticism I'll cite is less about what the phrase "data about data" leaves out and more about its imprecision relative to real-world examples. In a 1977 U.S. government manual on information management, for instance, the authors qualify their recommendations with this warning:

> It would have been awkward, and in some cases, almost impossible, to label each reference of information as to whether we were talking about substantive information or whether we were talking about metadata. The reader should, in the context of the discussion, determine which of these two types we are talking about, if, in fact, we are not talking about *both* substantive information and metadata. (Commission on Federal Paperwork, 1977, p. v, emphasis in original)

In the eyes of some later critics, this difficulty in drawing the line between data and metadata—between the "substantive" and the secondary, the content and the context—is more than reason for caution; it is reason for near-total abandonment. Few state the case as clearly as open data activist Michael Kreil (Exposing the Invisible, 2016):

> I have a problem with the term "metadata." I don't think that this term is precise, because, simply put, the basic idea of metadata is that it's data about data. For example, if

> I take a photo, I can add data like the camera model, time and geolocation, so, the additional information about when and where the photo was shot is called metadata. But, for example, if I take a lot of photos, I can use the metadata contained in these photos to connect the location in which I took them with the time I took them. The metadata can be used to track me. So, from that point of view, metadata is the data itself, and that's the interesting aspect, not the photos themselves.

For Kreil (and it's important to note he's not alone in this sentiment), the upshot of the term's imprecision is not a recognition that the prevailing conception of metadata is too limited. It is a recognition that metadata is rarely just metadata and that therefore the use of the term should be quite restricted. "In general, from a public point of view," Kreil clarifies, "everything is data, which is usually about persons. So let's stop calling it metadata." If some people continue to call things metadata, these should only be people who "add data to the data."

Fluid, Multiple, Fractional, Non-Neutral

"Data about data" is the most efficient and inclusive definition of metadata; it is misleading; it is restrictive; it is imprecise. Taking stock of these four takes on the dominant definition (and surely there are others), we find ourselves at a conceptual crossroads.

At first blush, it seems reasonable to take the path that has become familiar, the one that accords with prevailing cultures of metadata and that is arguably implicit in each of the above arguments. We assess which string of words can serve as the best possible standard definition, the one that will hold up for a variety of users across a variety of contexts, a definition that can encompass the "complex technical and intellectual infrastructure to manage and retrieve digital objects in different digital contexts, within different digital information systems and services" (Mendéz and van Hooland, 2014). We can ask: Is it better to open the concept of metadata to a wider set of objects, or is it better to keep things quite restricted, as Kreil suggests? If it is better to open things up, is the Pomerantz emphasis on informative statements the most effective and inclusive, or is even that approach beset by problems of restriction and bias, erroneously assuming it's all about information and underplaying metadata's social and material entanglements? Would it make more sense to accept the aphorism, then develop more precise definitions for specific

fields? Or should we just forget the whole business of questioning core concepts of metadata, as though this was a case of solutions in search of problems?

As compelling as these questions are, they prove unreliable when subjected to more critically and historically intensive scrutiny. It's convenient to treat metadata as a "definite and singular" concept, to use Matthew Mayernik's helpful framing (2020). One assumes that when Philip R. Bagley (apparently) coined the term in the 1960s, he succeeded in identifying a new element within the world of data and information; that element would grow and proliferate, but the core structure—something like data or information that describes other data or information—would remain the same. And yet this isn't what happened. For one thing, as Koster's insistence on the social and material reach of metadata suggests, there has been a significant increase in the variety of metadata types. Mayernik notes that metadata are as varied as "file naming conventions, catalog records, data descriptions in repositories, user tags on YouTube, notes in personal Excel spreadsheets, email headers, and HTML tags." To this list we must add further varieties in the realm of "user tags": forms of metadata that involve the passions and perspectives of everyday people, such as the hashtags that serve as the slogans and calling cards of social movements. We can also add more than types. Mayernik notes that people perform metadata "differently in different social settings and situations." They enact metadata as "Dublin Core records created by information professionals"; they enact it as "descriptions in lab notebooks created by scientists to document their data"; or (I would add) they put them into practice as part of the language of social media, as with the slew of metadata elements that have become second nature to those who frequent platforms like TikTok: the likes, hashtags, timestamps, usernames, and even the textual labels overlaid onto videos, the ingredients of a media practice I think of as "expressive folksonomy" (Parry, 2023).

At least one other development undermines the idea of metadata as enduringly definite and singular: it has become increasingly clear that metadata cannot be separated from economic and political domination. This is evident in the writings I've cited, as when Kreil refers to the uses of metadata as data for surveillance, or when Pomerantz recalls the use of metadata as the basis for killings by the United States military, as voiced in former CIA and NSA director Michael Hayden's unforgettable admission, "We kill people based on metadata." Metadata's relationships

to domination are also evident from other angles, such as recent projects in reparative and decolonial description and knowledge organization. Here there is a recognition that countless descriptive inheritances—terms, summaries, inventories, taxonomies—owe their origins and outlines to histories of racism and imperialism, and that conscious efforts must be made to assert other vocabularies, other understandings, other voices, and other stakes, in however limited or imperfect a fashion (Adler, 2016; Frick and Proffitt, 2022). And this is just one among many power-conscious and justice-driven approaches to metadata. There are, for instance, the efforts of indigenous communities to contest only seemingly universal protocols like Dublin Core through Traditional Knowledge (TK) labels, as explored by María Montenegro (2019). There is also Thomas N. Cooke's reframing of the seeming marginality and banality of metadata as a smokescreen, as with the "coalescing of mutual interests within the big data and smartphone manufacturing industries to maintain the invisibility and illegibility of metadata" (2020, p. 92)—metadata as a murky "mechanism of capital," a kind of half-open, sociotechnical secret.

Taking all these developments into account, one can no longer treat a list of basic conceptions of metadata as a straightforward menu of options. Not only are we faced with a notion that didn't come fully formed—Bagley notably put the word "about" in quotation marks when introducing the neologism (Bagley, 1968, p. 26)—we also face a set of multiple, mutually compelling propositions with respect to how that term should now be defined, conceived, and enacted, including the idea that the term should be effectively discarded in favor of full-fledged attention to data. We are also (or so one hopes) confronted with the need for both relative clarity on the objects of concern and a meaningful orientation toward the fraught and exploitative dimensions of the phenomenon as it exists and is put to use, no longer assuming that metadata is some neutral substrate or afterthought, instead heeding its evolving roles in practices of extraction and domination.

The question is what to do. My fundamental argument in this chapter is that the time has come to test new approaches. We can start by taking as given what Mayernik lays out: that metadata is not "definite and singular" but "fluid, multiple, and fractional," which is not to say it is infinitely redefinable, but that it is marked by conceptions and sub-conceptions that are many, interconnected, and at times contradictory. We can then add to this list of traits, asserting that the

concept is also non-neutral; similar to how systems of classification "impose their own rationale" and assert a "way of seeing the world" (Montenegro, 2019, p. 737), conceptions of metadata favor certain ways of thinking, seeing, and valuing over others. Having adopted these critical frames, we can then proceed on a path of creative skepticism. We can suggest that the conceptual landscape of metadata is insufficiently elaborated, and we can then ask what alternative definitions, categories, and typologies will prove meaningful and useful. None of this would be done in the spirit of total correction because that would be out of step with the basic premise of metadata as a fluid, multiple, fractional, and non-neutral concept. Rather, this would be done in the spirit of plurality and justice: metadata understood not from the perspective of orthodox objects like canonical paintings and commercial barcodes, but from the perspective of contrasting and contested practices, whether that's the exploitation of machine-learning assessed user characteristics or anticolonial tagging.

In short, to ask why and how metadata isn't (or isn't only) "data about data" is to see "Who's right?" in one direction and "What else?" in another. Taking this latter path, we pose questions about metadata that combine the conceptual, the critical, and the creative. What has been left out, overlooked, or overemphasized in conceptions of metadata, and how might these silences, absences, and imbalances be meaningfully counteracted? Is there reason to abandon the concept altogether, as Koster and Kreil suggest, or is it possible to take a reparative or even radical approach to metadata?

Data Near Data

Of the many conceivable ways into this terrain, testing hidden premises is the one I am best positioned to attempt. What's something that is consistent across both the default and the corrective definitions that might yet be called into question? How might this consistent element be rethought with the help of fields not typically understood as concerned with metadata?

Most obviously, there are the consistent elements of data and information, but there is also that other term Pomerantz briefly addresses: "about." With the caveats that no definition is neutral and that aboutness remains to be theorized in depth, we can lean on the definition offered by

Stephen Yablo (2014, p. 1) in an eponymous book: aboutness is "the relation that meaningful items bear to whatever it is that they are on or of or that they address or concern." To say that this relation is prevalent in the world of metadata would be an understatement. Not only did Bagley's framing of the term put aboutness front and center, saying that the type of data for which he didn't quite have a name could be thought of as data with the quality of being "about" other data, subsequent discussions of metadata in the world of libraries have often included discussions of the subject matter of texts, and subject matter is often the thing we speak to when we use this preposition (Holley and Joudrey, 2021).

Adding to the apparent appeal of aboutness, it seems to hold up quite well when applied to new metadata varieties. For instance, in adding a description like "the problem is capitalism" to an image of a factory floor posted to social media, a user would typically be indicating that the post is somehow related to or on or from debates on political economy; the description tells us something about the post's subject matter. At the same time, another type of metadata that attends the post, the number of accumulated likes, is not an indication that the post concerns capitalism and other political economic forms; it is a potentially informative statement about the status of the post in comparison to others. Whether or not users take the time to read these different metadata elements, they will find themselves confronted with multiple informational relations at the nexus of metadata and aboutness. That nexus isn't only significant for users: the post also serves as a convergence of tags, topics, and use data for businesses. For example, the fact of a user has taken time to pause and read the post could be treated as indicative, adding to the trove of data about them (and about those deemed to share their categories) and potentially altering the kinds of topics assumed to interest them for the next refreshes of their feeds and "for you" pages, even serving as avenues for their continued capture in loops of attention and extraction (Cooke, 2020). In other words, what is explicitly a post about (concerning, of addressing) capitalism is also, structurally speaking, a post "about" (having to do with, turning on, relevant to, enmeshed within) surveillance capitalism (Zuboff, 2019).

Given these ways in which aboutness is not only quite ingrained but also persistently active and germane, it's clearly going to require some shaking up of our habits and assumptions to see anything else at work, and there's no guarantee that alternatives will be valuable. One thing we

can do is zoom out to consider ideas and approaches from fields not traditionally concerned with metadata, such as visual studies and cultural theory. More importantly, we can do something this collection encourages and which I, as a scholar who tends to write about culture and media in the United States, am learning to do by way of unlearning: we can also look to writers who have been critiquing Western epistemologies for a long time. What these writers often do is refuse to take fundamental assumptions as givens, instead seeing that the concepts that dominate European and Anglo-American discourses come from particular places, produce oversights, and serve particular purposes, including purposes of exploitation and domination (Smith, 2021). Crucially, these writers will not look at the prevalence and usefulness of a framework like aboutness as any assurance of an essential or universal status. Instead, as is already the case with a small set of contemporary artists, they will approach aboutness with attention to power relations. Aboutness can name the habit of fashioning a strict, stable argument or zeroing in on a relevant and topical subject matter (Berwald, 2016; Raina & Smith, 2018). It can be the posture of distanced description epitomized in imperial documentation of colonized peoples. Aboutness can even be reframed as a widely active construct that must undergo decolonization; it maintains the political, economic, and ecological status quo even under the veneer of urgency (Ashbel, 2019; Wenner, 2021).

Especially useful along these lines is a critical gesture that many writers have cited, and that I've cited elsewhere as well. At some point in the development of her 1982 film *Reassemblage,* Trinh T. Minh-ha determined that the dominant, aboutness-centered way of thinking about the project wasn't appropriate. One might think that a film made up of footage of women's lives in rural Senegal would be a film speaking "about" those women. But Trinh voiced a different approach in her voice-over; the film would be an exercise in "speaking nearby" (Chen, 1992). This means we pay attention to the location of the person and the text doing the speaking, both their geographic locations and their positions with respect to their living subjects (the people "in" the film) and the film's subject matters (what it addresses); in Trinh's words, this mode of speaking "reflects on itself and can come very close to a subject without, however, seizing or claiming it" (p. 87). To speak nearby is to attempt a kind of expressive proximity or even semantic allyship: a willingness to stand "with" those near whom one speaks while also trying to avoid a posture of objectivity.

Let me start by emphasizing the considerable distance that sits between Trinh's efforts at decolonizing documentary in the early 1980s and a reexamination of metadata conceptualizations in the early 2020s. It's important to keep this distance in mind, recognizing key differences in purpose, context, and medium, among others. And yet that doesn't mean we ought not put them in conversation. Here the generative overlap is that crucial element of world languages: prepositions. These are short words (also called adpositions and not always words per se) that fly under the radar, all the while doing essential communicative work (Kurzon and Alder, 2008; Hagège, 2010). Speaking of English in particular—and not pretending that English can serve as a stand-in for all languages—the linguist R. M. W. Dixon says that prepositions "play a vital role in the language": they "indicate how and where, when and why, purpose and association, inclusion, connection, and many other things" (Dixon, 2022, p. vii).

Dixon's list is apt for considering prepositions as they bear upon the current discussion. Trinh saw in the use of "about" a mode of distanced description that maintains a veneer of neutrality. She then engaged in a practice of prepositional substitution, a practice that opened other kinds of purposes, associations, inclusions, and connections in filmmaking. Prepositional substitution in the case of metadata conceptualizations is both similar and different. What is similar is the questioning of "about" as the seemingly obvious and objective preposition to describe what metadata is, are, and do. (It's worth noting that "data" and "metadata" are used as both singular or plural nouns; I tend to employ the singular for readability.) What is different is how we respond. We aren't necessarily looking for a different way of making and managing metadata per se, although that is certainly a path of inquiry that can be and has been followed, as with feminist and anticolonial archival practices in which the "archivist cares about and for and with subjects" (Caswell and Cifor, 2016, p. 36), or what Itza A. Carbajal (2021, p. 103) calls a "communal metadata praxis." Rather, we are looking for ways of addressing metadata in which aboutness is no longer the singular genetic element or the boundary line, when other versions and instantiations of metadata are analyzed and reworked. In short, we test prepositional substitution as a tool for remaking the landscape of metadata conceptualizations, and we do so with attention to the past, present, and future of alternative worlds—those not defined by the hierarchies and pathologies of capitalism and colonialism.

Here a simple phrase comes to mind: "data near data." Whereas the preposition "about" is only sometimes spatial, tending to refer to situations and matters of concern, the preposition "near" is specifically spatial, carrying the basic meaning of "in the vicinity of." The term can also be "extended to cover physical and mental conditions," as in "Agatha was near despair all winter" (Dixon, 2022, p. 326). Applying that definition to metadata in an unfiltered way, we get phrases that echoes Yanni Louikissas's mantra that "all data are local" (2019): data in the vicinity of data, information close to objects, a potentially informative statement alongside a potentially informative object, even data and information near people. To judge whether these are useful provocations is to judge whether they meaningfully answer this chapter's driving question with respect to what has been left out, overlooked, or overemphasized in conceptions of metadata.

Let's start by aligning several of the examples of metadata I have mentioned in this chapter:

(a) Social metadata. Across various platforms, there are fields for the display of metadata that depend on user contributions. These can be quantitative, such as like counts. They can also be qualitative, such as additions in the caption or description field.

(b) Hashtags. According to Elizabeth Losh, a hashtag is a type of metadata tag that "groups similarly tagged messages or allows an electronic search to return all messages that contain the same hashtag." Whether accompanied by a "word, abbreviation, acronym, or unspaced phrase," the hash character "tells the computer that a particular word or words should be read as more important than other words in a given message for purposes of sorting digital content into similar clusters" (Losh, 2020, p. 2).

(c) TK labels. As Montenegro (2019, p. 739) summarizes, TK (traditional knowledge) labels are "a set of 17 digital tags that can be included as associated metadata into diverse digital information contexts—CMSs, online catalogs and databases, finding aids, online platforms—assisting in the recognition of, and education about, culturally appropriate circulation, access and use of Indigenous cultural materials." As part of an "extra-legal initiative," these tags "draw from and extend already existing community protocols as their base," and they specifically diverge from universal standards, such as Dublin Core,

that, as Montenegro shows, are incommensurate with many Indigenous understandings, uses, and protocols for cultural materials.

Now, to look at these examples from the perspective of "data about data" is to see the throughlines that have tended to inform a great deal of metadata discourse: there are forms of secondary data; these data do different kinds of description; and that work of description supports storage, access, and retrieval.

To look at these examples from the perspective of "data near data" and its correlates is quite different. For a start, we see that we are not in the territory of essences and either/ors. That's not to say the different examples are secretly all about vicinity. It's to say that there is the potential for multiple kinds of relations. Metadata might be a field of, as it were, prepositional and relational abundance.

Adopting this pluralistic way of looking, we get to multiple versions of what nearness can mean in the context of metadata. Most basic (but not for that reason any less important) is a variable I'll call *visible position.* A significant number of metadata types and metadata examples derive part of their meaning and function from being nearby. That nearby location can be quite fixed. For instance, the variations on social metadata that attend posts are located where the app design beckons they be placed. The location can be modestly variable. A user can decide where to place hashtags and in what order, whether within the body of a description or in a collection that follows a description. And the location can also be a matter of interruption. When TK labels are affixed to an entry in a database, they are there to interrupt an otherwise established view. They are there to be read and understood for different ways of understanding and interacting with that object. The TK labels assert a position that would not otherwise have existed: a place for community protocols.

All of this could sound like a mistake of categories, as though we were talking about expression, interpretation, and use when we should be talking about management and access. It is at this juncture that premises within the study of visual culture prove helpful. Losh says that "critics of visual culture assume viewers are inculcated into seeing specific things, and they are also trained

to block out things and render them invisible. In this way everything from an evidence photograph to a medical scan to a frame-grab of online video is seen through the lens of prior education that disciplines how eyes scan the visual field" (p. 53). For Losh, hashtags are not somehow observers of or even just side players in these processes of visual discipline. Hashtags likewise "train the eye and set expectations about what to see" (p. 54). (She gives the example of materials tagged with #Love during the variously named 2014 Maidan Revolution in Ukraine. "The expectation for these images," writes Losh, "would be that they offer a vision of cooperation rather than conflict, such as shots of egalitarian interactions that convey the positive emotions associated with a leaderless revolution in which everyone participates.") Losh does not say that this dimension of hashtags is a dimension of nearness, but it is arguably implicit in the account. After all, these hashtags couldn't do this work—and TK labels couldn't do their interruptive pedagogy—without being near and readable. Not only that, these proximate metadata are indications of other kinds of mental and conceptual locations: positions and perspectives the viewer/reader/user can see the content as addressed to, or apparent places within hierarchies of interest, popularity, or other dimensions of which the post is representative.

The latter notion suggests another variable at the intersection of metadata and nearness: what we might call *tactical approximation.* Consider this advice from *Popular Science* (any number of similar examples could have been selected)*:* "If you need inspiration for finding related hashtags (or coming up with more to add to your own videos), try the Hashtag Generator for TikTok online, where you can type the name of a label and the platform will show you a list of similar ones you can use and tell you how popular they are" (Nield, 2022). Should people on TikTok follow these instructions, they will be exploring and considering (and having a database and algorithm tell them) what is in the relative referential orbit of their video and proposed hashtags. Some users might take the work of figuring out proximate hashtags as an interpretive or even expressive exercise, concluding, for instance, that #capitalism should sit alongside a video on climate, or that the issue of #extraction is worth asserting as near a discussion of a recent report on museum commissioning practices. At the same time, as the advice I've quoted indicates, there is no purity of purpose in figuring out the strings of metadata that are potentially proximate to social media content. Metadata choices can easily be just a matter of present popularity: appending this hashtag to this content is an attempt at expanding the post's circulation, as with

adding #cute to a post that is nothing of the sort. Aboutness is certainly in play here, but that term tends to privilege frames of truth and falsity. When we look at the full spectrum of meta elements, including those that are highly structured, those frames don't hold. Metadata producers often value the sufficiently close and the tactically proximate. That can be the case when content creators labor over their metadata, hoping to put their materials and handles in phones and eyeballs and minds. (Content creators are also metadata creators.) It can also be the case when cataloguers engaged in reparative description revise the false and harmful categorizations and summaries of the past, replacing these false and harmful approximations with better, feasible, and yet also imperfect approximations of just meta information in the present.

Of course, it isn't only people who produce, position, and read metadata; these are also things done by various profit-driven technologies—machines, software, and algorithms—and this gets us to a third version of metadata through the lens of nearness, what we can call *hidden vicinity*. The messaging app WhatsApp provides an especially helpful example. The straight story of the platform is that the messages you send are end-to-end encrypted, meaning they can't be read by the company; they can only be read by the people who have the encryption keys. And yet the metadata that accompanies and is produced by these exchanges is not encrypted. The company can track and share (does track and share) with whom you message, when, and how, and these forms of metadata can coincide with or be integrated with other forms of computational metadata more difficult to trace and summarize, as Cooke (2020) lays out. Meredith Whitaker, the current president of Signal, a messaging app that does encrypt its metadata, explains: WhatsApp "collects the information about your profile, your profile photo, who is talking to whom, who is a group member. That is powerful metadata. It is particularly powerful—and this is where we have to back out into a structural argument—for a company to collect the data that is also owned by Meta/Facebook. Facebook has a huge amount, just unspeakable volumes, of intimate information about billions of people across the globe. It is not trivial to point out that WhatsApp metadata could easily be joined with Facebook data, and that it could easily reveal extremely intimate information about people" (Patel, 2022).

In a sense, I have already introduced the phenomenon this example suggests: I have spoken to the ways that social media posts serve as avenues for businesses to learn something about the

user or the categories of people they're deemed to occupy (however unreliably, however cynically). What I have left out, however, are the dual dimensions of proximity and obscurity in all this. This is a complex terrain to navigate—that's part of the point—but the label "hidden vicinity" might sufficiently condense the dynamics taking place. For one thing, there is the sense in which the metadata extraction is not advertised or indicated as such, unlike the encryption of the content; it is hidden and even purposely obscured, as Cooke (2020) suggests. Then, there is the fact that this metadata extraction is occurring right alongside the interaction; it is a feature lurking with and alongside the practice of speaking. Then, there is what coincides with the extraction, as Zuboff (2019) and many others have spoken to. The extraction from the obscured environment serves injections of material into the visible environment. The content you see, the suggestions you receive, the advertisements that appear, when and how these things occur—all of this is a matter of what can find its way near you, close enough that it can be received and acted on. In other words, the cycles of interaction—content, algorithm, data, metadata, analysis, delivery—are implicitly geared toward keeping one close, keeping one coming back, shaping a subject who is near their phone, near their devices, in the vicinity of a potential purchase, or a potential vote. Proximity is power.

In stark contrast to hidden vicinity is one final variation on metadata and nearness I'll emphasize; it is a variation I suggest is embodied in TK labels. One thing they seem to say on aggregate harkens back to both Trinh's intervention and those of the artists who see in aboutness the striated, extractive ways of seeing and collecting that distinguish colonial modernity. TK labels suggest a way of conceiving and understanding metadata that is not centered in the speed and generality and consistency of knowing (and having the implicit, even "imperial right" to assert) what anything in the world is about (Azoulay, 2019), but that is instead centered in the caution and texture and relationality of knowing when and how to interact with what is, for the moment, near, or of knowing when something ought not, or ought not yet, be near. If the implicit paradigm of metadata has been access on demand for those in a position to act on that access, TK labels suggest a different paradigm, one that might be tentatively labeled *careful and contingent proximity.* Montenegro's read of TK labels as a "form of contestation" is especially apt here. At one level, she shows the concrete ways these labels insist on careful and contingent relation and engagement. For example, the Seasonal label, used by the Pokagon Band of Potawatomi Indians,

serves to "indicate that the circulating material should only be heard and/or utilized at a particular time of the year" (p. 740). As another example, the Secret/Sacred label has been used by the Sq'éwlets people to let "external users know that the material that is circulating should not be publicly available due to its secret, sacred, or esoteric components" (p. 741). At another level, Montenegro shows how the use of the labels exceeds any particular function. She says that the labels "broadly convey Indigenous concerns around access, circulation, and use," and that they eschew use as "a fixed application profile within a particular standard." Instead, they "work as a mechanism that refuses to erase the hegemonic condition of standard metadata fields, leaving those exclusionary terms visible and evident"—keeping these terms nearby—while also contributing to "the enactment of Indigenous peoples' inherent right to control and govern the gathering, management, circulation, access, use and interpretation of their own data, thus contributing to tribal self-determination, data governance and, ultimately, sovereignty" (p. 743).

This question of appearance, circulation, and care implies another prepositional strand. In her account of hashtags through the lens of visual culture, Losh doesn't only emphasize the level of the individual person or image. She also points to the work of hashtags at the wider scales of aggregates of images, or even the broader and more diffuse phenomenon Jacques Rancière labels the "distribution of the sensible." "In many ways hashtags are a mechanism for the redistribution of the sensible," Losh (2020, p. 56) observes, "by making particular kinds of content more perceptible to potential audiences . . . The act of labeling a chunk of data shines a spotlight on it for a heterogeneous audience that may be composed of both activists seeking to reach a critical mass of participants and security forces planning their crackdown on a crowd of unruly dissidents." In other words, the label or labels that are added to a given "chunk of data" are not just in an informative relationship with that chunk (about), nor are they also posited as somehow expressively or tactically proximate (near); they are also there in an active and supporting role, serving to facilitating that chunk's appearance or persistence in the spotlight or, to speak near TK labels, those materials' considered and accepted passage in and out of circulation.

The preposition that comes to mind for these relational dynamics is "for"; these hashtags could be understood as acting "on behalf." Part of the facilitating work they do is to aid in the distribution (or non-distribution) of what is considered the central "content." Further work is for

audiences. The hashtag "says" something informative about an event or an idea, but it supports the apparent informational needs of one who would care about this topic. Without it, the participants might not see the content in question; the metadata counters a potential scarcity or missed opportunity. Of course, as Losh emphasizes, hashtags can also work on behalf of security forces looking to quell dissent, or they can serve as reservoirs of data about people for the larger machines of attention and engagement. And none of this is restricted to hashtags: many contributions to social metadata carry with them metadata elements for which the operations "for" are active and real, even if harder to precisely quantify or summarize. The difficulty of naming and analyzing certain meta relations doesn't make them any less significant.

From "about" to "near" to "for"—as I hope is obvious, much more could be said about the two additions, and the list of prepositions can easily grow. (For instance, hashtags suggest metadata can exist "with" or even "in" other data and information, and we also need to consider relations like "against" and "from.") But the more pressing concern for present purposes is whether the provocation "data near data" has meaningfully altered our conceptual purchase on metadata.

I contend that it has. Not only have we managed to evade what we might hazard calling the "hegemony of aboutness," we have also been able to see one further reason that metadata is not just data about data. Whether as individual items, aggregates, or even taxonomies and ontologies, metadata bear more varied and complex prepositional relations than the normative frame of aboutness lets on. My contention is that those prepositional relations exist and matter. My further contention is that the slogan data about data occludes rather than illuminates them. The critiques of the phrase get us some way there, but it is necessary to go further—to dispense with the non-natural and non-neutral assumption that it's all about aboutness. Metadata is far more abundant in social and expressive relations than both the default and corrective definitions would indicate, and "data near data" is an effective, if necessarily imperfect, initial corrective, one that seeks to place decolonial, anticapitalist, and other intersecting critical perspectives at the heart of the discussion.

Metadata Beyond Aboutness

Many questions remain, such as how different meta and prepositional relations interact, or how to cultivate more cultural and linguistic specificity with respect to both literal prepositions and adpositions and their analogues in metadata, or whether and how to extend the concept of metadata to pre-computational and non-machine-readable creations (Weinberger, 2007). Rather than address such questions all too quickly, I'll close by returning to those that have driven this chapter. What might be done differently in inquiry and practice when prepositional abundance is a given? And how should the landscape of metadata conceptualizations be altered to accommodate the perspectives I've explored here?

Taking a cue from metadata's fluidity and ambivalence, I'll answer not just one way but several. One possibility is that we could basically slightly append the standing approach to say that metadata is effectively data about* data or a potentially informative statement about* a potentially informative object, with the asterisks indicating the inevitable presence of other prepositional modes. The appeal of this approach is its mix of simplicity and open-endedness; the asterisk can hold space for any number of reasons why aboutness should be qualified, from Pomerantz's emphasis on subjectivity to much more involved and indeed politically informed critiques, including those that frame aboutness as a fundamentally colonial construct. And yet it seems that these affordances come at a cost. For one thing, the use of about* leaves the frame of aboutness firmly intact, a status that is open to objection along not just the lines of precision but also of value and power. The use of the asterisk also risks confining interpretation and expression to the margins of debate, allowing the discussion to revolve around managing objects and information, when the kinds of examples I've assembled here suggest the vibrancy in expressive meta statements and performative data utterances, and when scores of people regularly engage with metadata in exactly these terms. Finally, the use of the asterisk doesn't get at the dynamics of power I identified above. It does nothing to indicate the non-neutrality of the concept, nor the role metadata might play in repair, nor the potential for new conceptions (or refusals) of the phenomenon from decolonial, anticolonial, and antiracist positions.

A more thorough embrace of the implications of prepositional substitution leads us to more intensive proposals. One option is to build on Koster and Kreil's provocation that we ought to effectively stop using the term metadata. We could say that metadata in the context of social

media and widespread surveillance reveals all the trouble with the term metadata: that hashtags can count, with all their varieties of functions, would seem to reveal how confusing and superfluous the original concept always was, and how it's much better to frame everything out there as data. Taking this position further, we would get to the idea that there are many types of data and information in the world, and that hashtags and other things we call metadata are better understood as pieces and sets of information with multiple prepositional functions and switches built into them. To that end, we could coin a term that would subtly but meaningfully distinguish particular types of data produced by people and machines: "prepositional data." These would be data that serve to indicate and mediate "how and where, when and why, purpose and association, inclusion, connection, and many other things." I can imagine further study of both individual prepositions and their interrelationships. I can also imagine analysis of how prepositional data can shed that distinguishing prepositional feature, coming to be treated as "just data," with this switch of status often taking place in the shadows.

Another option—the last I'll indicate—is to move away from the focus on definitions. After all, there is another, equally significant dimension of the landscape of metadata conceptualizations. Mayernik (2020) speaks to it: "If the notion of 'metadata' has been defined in a variety of ways, it has been categorized and conceptualized in an even more diverse fashion. Categorizations of metadata reflect the different conceptions and motivations of the people who generate them, and manifest in a variety of metadata typologies." Mayernik gives many examples of these categorizations, such as the "constructed, constructive, actionable" triad, or the five types laid out by Gilliland (2008): administrative, descriptive, preservation, use, and technical. This second set is especially relevant to the questions I've raised here; it is the kind of typology that tends to accompany "data about data," whether this is in lectures, textbooks, explanatory videos, or the results of web searches. The typology acts as an entrance point and initial filter for many people first introduced to metadata; it also structures projects and even professions.

From the perspective of this chapter, an intervention at the level of categories suggests additions rather than deletions. Several are already on the table. When institutional actors engage in the work of altering or newly developing metadata toward antiracist and anticolonial ends, they can be said to be in the practice of thinking, producing, and enacting *reparative metadata*, and that

category suggests other crucial kinds, such as *decolonial* and *anticolonial metadata.* A further addition was implicit in my discussion of nearness. As we have seen, hashtags can certainly function in the mode of descriptive metadata, but quite often they function in a mode that is closer to speech. They serve as statements in themselves, or the statements they make in themselves come into dialogue with the notions or feelings or ideas in what is otherwise considered the primary material. This is a version of metadata we can dub *expressive metadata*. Here "expressive" refers to the work of mediating—conveying, showing, assembling—thoughts, feelings, and ideas. A further addition was implicit in the discussion of "forness": *distributive metadata*. These would be metadata that help distribute pieces of content through informational networks; they support that content's spread, though they can also be as much "about" circulating themselves as they are about circulating given content. This is evident in political hashtags that assert a norm or a value, but it is also the case with other hashtags. These become complex entities in themselves, entities that manage to maintain or expand their distribution their yoking to other materials and other metadata.

If there is a potential downside to these additions to the core typology of metadata, it is that they do not necessarily answer all the lines of critique of data about data. For example, the concerns of Pomerantz and Kreil persist: the seeming need to maintain the frame of information, or the concern that, at this point in the game, it's all just data and mostly data about people. One potential answer to these doubts is to take the inclusion of "expressive" metadata seriously. We can see in the rise of expressive metadata the implication that the world of metadata—the name notwithstanding—isn't strictly a world of data. We might even venture to think of metadata as a cultural form for which data and information have historically been the dominant frames and modes, but for which other frames and modes are possible, including the layers of expression, the positing of relationships, metadata as a "compilation of words used to express ideas, feelings, emotions, and values" (Carbajal, 2021: 102). Some of the most powerful uses of social metadata—the kinds that have helped spark protests, change minds, move policies—have been those that take up this meaning-attentive approach. They creatively combine metadata types: the reparative with the descriptive, the distributive with the expressive. At the same time, as many of the writers I have cited in these pages would emphasize, it remains to be seen whether these

critical and intersectional uses of metadata can overcome and outlast that most persistent of metadata types: the extractive.

References

Adler, M. (2016). "The Case for Taxonomic Reparations." *Knowledge Organization* 43(8): 630–640.

Ashbel, A. E. (2019, August). "An Argument against Aboutness." *Vorherbst Magazine.* https://www.steirischerherbst.at/de/vorherbst/591/an-argument-against-aboutness.

Azoulay. A. A. (2019). *Potential History: Unlearning Imperialism.* Verso.

Bagley, P. R. (1968). "Extension of Programming Language Concepts." University City Science Center.

Berwald, O. (2016, Spring). "Against Aboutness: The Corpse that Almost Got Away." *The Scofield*, 1.4, 196–197.

Carbajal, I. A. (2021). "Historical Metadata Debt: Confronting Colonial and Racist Legacies through a Post-custodial Metadata Praxis." *Across the Disciplines*, *18*(1-2), 91–107.

Caswell, M & Cifor, M. (2016). "From human rights to feminist ethics: Radical empathy in the archives." *Archivaria*, 23–43.

Chen, Nancy N. (1992). "'Speaking Nearby': A Conversation with Trinh T. Minh-ha." *Visual Anthropology Review* 8, no. 1: 82–91.

Commission on Federal Paperwork (1977). *Reference Manual for Program and Information Officials: A Handbook for Managers.* Volume 1 of a 2 Part Manual.

Cooke, T. N. (2020). "Metadata, Jailbreaking, and the Cybernetic Governmentality of iOS: Or, the Need to Distinguish Digital Privacy from digital privacy." *Surveillance & Society* 18(1): 90–103.

Dixon, R. M. W. (2022). *English Prepositions: Their Meanings and Uses.* Oxford University Press.

Exposing the Invisible (2016, May 11). "Michael Kreil: An Honest Picture of Metadata." https://exposingtheinvisible.org/en/articles/michael-kreil.

Frick, R. L. and M. Proffitt (2022). "Reimagine Descriptive Workflows: A Community-informed Agenda for Reparative and Inclusive Descriptive Practice." Dublin, OH: OCLC Research.

Furner, J. 2020. "Definitions of 'Metadata': A Brief Survey of International Standards." *Journal of the Association for Information Science and Technology* 71, no. 6: E33–E42.

Gilliland, A. J. 2008. "Setting the Stage". In *Introduction to Metadata: Pathways to Digital Information*, 3rd. ed. M. Baca. Getty Information Institute.

Hagège, C. (2010). *Adpositions.* Oxford University Press.

Holley, R. M. & Joudrey D. N. (2021). "Aboutness and Conceptual Analysis: A Review." *Cataloging & Classification Quarterly* 59:2-3, 159–185.

Koster, L. (2015, June 16). "Standard deviations in data modeling, mapping and manipulation." https://commonplace.net/2015/06/standard-deviations.

Kurzon, D. & Adler, S. *Adpositions: Pragmatic, Semantic and Syntactic Perspectives.* John Benjamins Publishing Company.

Losh, E. (2020). *Hashtag.* Bloomsbury Academic.

Loukissas, Y. A. (2019). *All Data Are Local: Thinking Critically in a Data-Driven Society.* MIT Press.

Mayernik, M. S. (2020). "Metadata." *Knowledge Organization* 47, no. 8: 696–713.

Méndez, E. and van Hooland, S. "Metadata Typology and Metadata Uses." In Miguel-Angel Sicilia (ed.), *Handbook of Metadata, Semantics and Ontologies* (pp. 9–39)*. World Scientific, 2014.

Montenegro, M. 2019. "Subverting the Universality of Metadata Standards: The TK Labels as a Tool to Promote Indigenous Data Sovereignty." *Journal of Documentation*, 75, no. 4: 731–749.

Nield, D. (2022, August 15). "7 tricks to make the most of TikTok." *Popular Science.* https://www.popsci.com/diy/tiktok-tips.
Parry, K. (2023). *A Theory of Assembly: From Museums to Memes.* University of Minnesota Press.

Patel, N. (2022). "Why Signal won't compromise on encryption, with president Meredith Whittaker." *The Verge.* https://www.theverge.com/23409716/signal-encryption-messaging-sms-meredith-whittaker-imessage-whatsapp-china.

Pomerantz, J. (2015). *Metadata.* MIT Press.

Raina, K. and Smith, J. (2018). "Q&A with Alice Tippit." *Maake Magazine.* https://www.maakemagazine.com/alice-tippit.

Ranger, J. (2012, September 19). "9 Reasons You Understand More About Media Preservation Than You Think." https://blog.weareavp.com/9-reasons-you-understand-more-about-media-preservation-than-you-think.

Smith, L. T. (2021). *Decolonizing Methodologies: Research and Indigenous Peoples* (Third Edition). Zed Books.

Weinberger, D. (2007). *Everything Is Miscellaneous: The Power of the New Digital Disorder.* Henry Holt and Company.

Wenner, S. (2021). "Aesthetics of Mykorrhiza. The Practice of *Apparatus.*" In C. Stalpaert et al. (eds.), *Performance and Posthumanism: Staging Prototypes of Composite Bodies* (pp. 93–109). Springer.

Yablo, S. (2014). *Aboutness.* Princeton University Press.

Zuboff, S. (2019). *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power.* Public Affairs.

Bio

Kyle Parry is Associate Professor of History of Art and Visual Culture at the University of California, Santa Cruz. He coedited *Ubiquity: Photography's Multitudes* (Leuven University Press, 2021) and is the author of *A Theory of Assembly: From Museums to Memes* (University of Minnesota Press, 2023).