Worked Up About Data

meditations on data work for the humanities

art by María Medem

Meredith Martin / Center for Digital Humanities at Princeton / DARIAH June 19, 2024 / mm4@princeton.edu / @mmvty.bsky.social

I dream about workflows

also I am very tired

I just wrote a book called

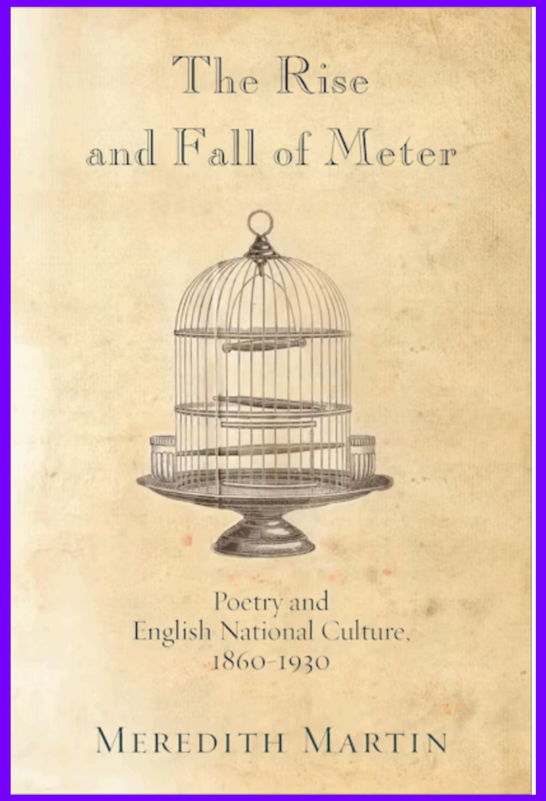
Poetry's Data

After this trip I will write

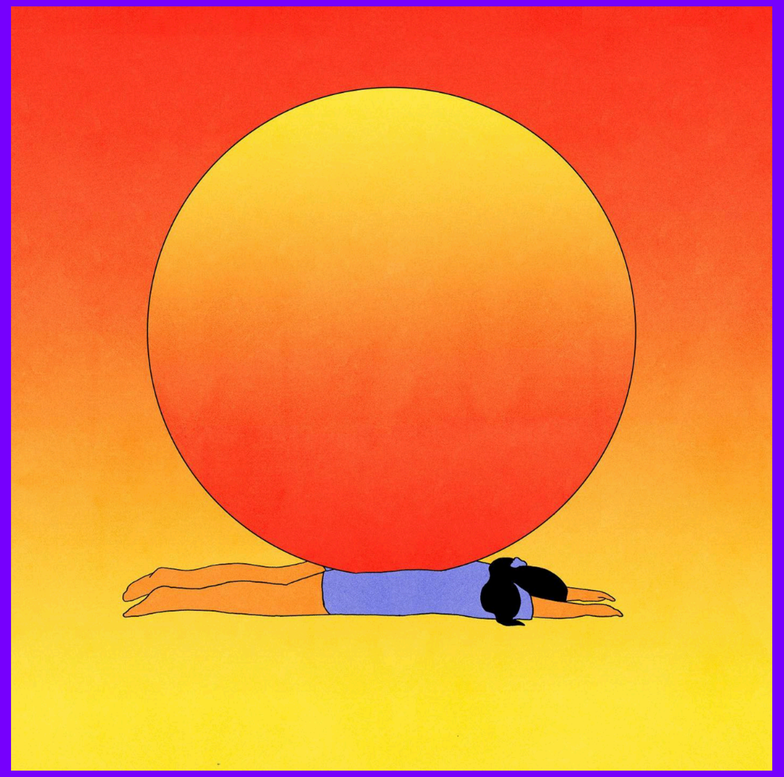
another book

Data Work in the Humanities

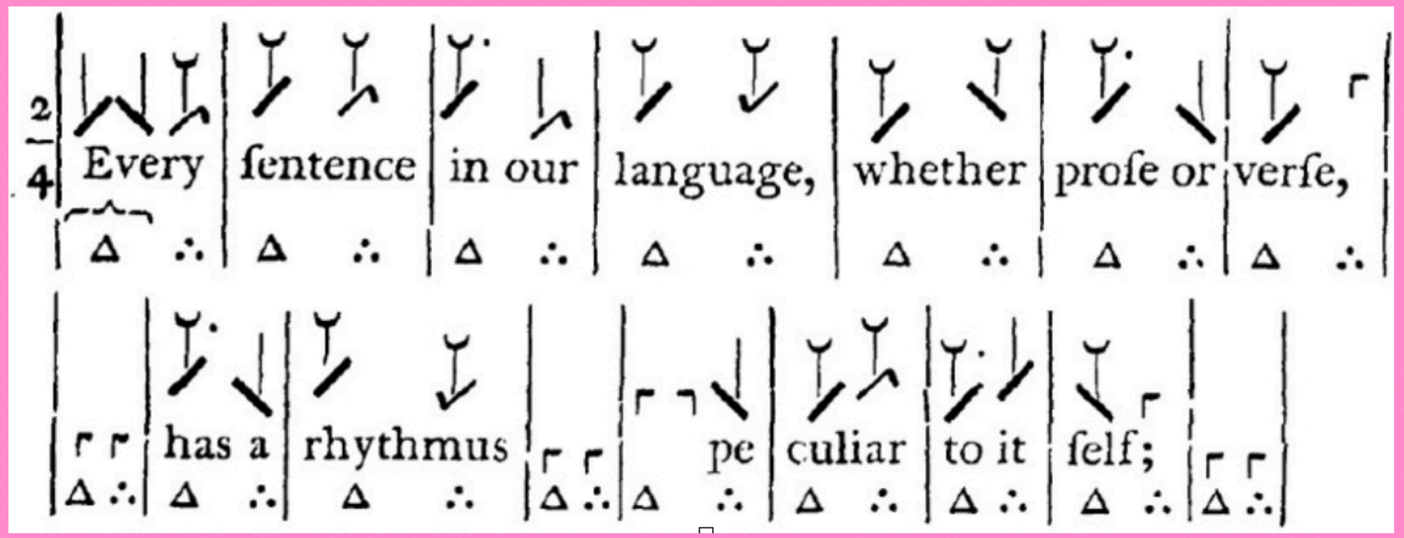
I have to write books. I'm a professor. In the U.S., despite all the changes in higher ed and despite the pace of technology, single-author monographs (and articles) are still seen as the highest — and often the only — form of knowledge production in the humanities.



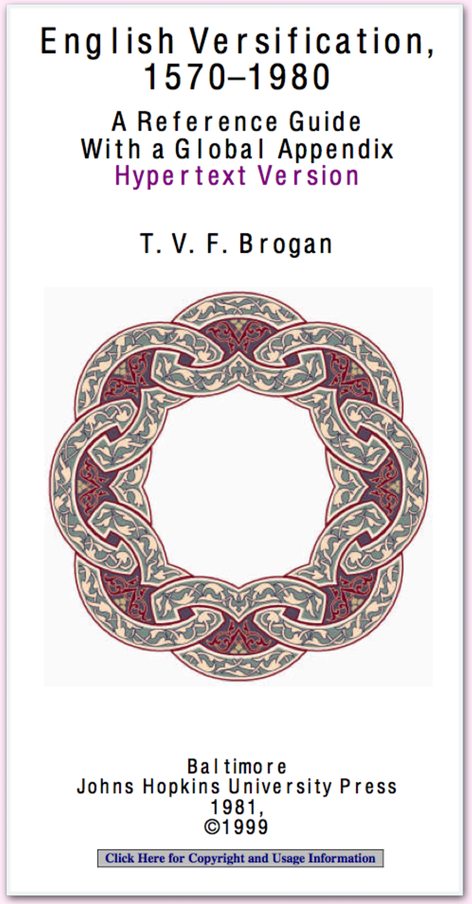
I like books.



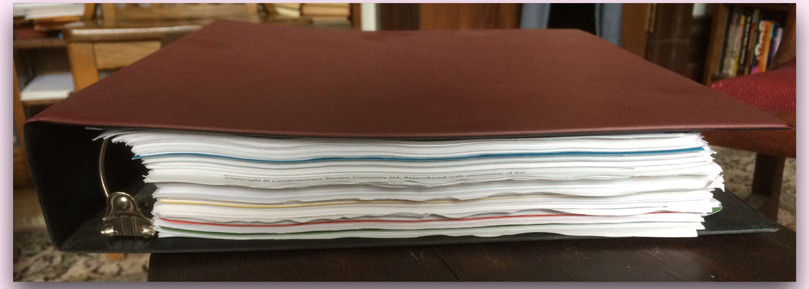
Poetry's Data argues that in our technologically mediated research environment, our old models of the autonomous single scholar no longer hold. We need to acknowledge both the embeddedness of our critical sources & research materials within multiple layers of mediation. And we must also acknowledge our own situatedness in humanities research ecosystem that demands collaboration and collectivity and interdisciplinarity to thrive.



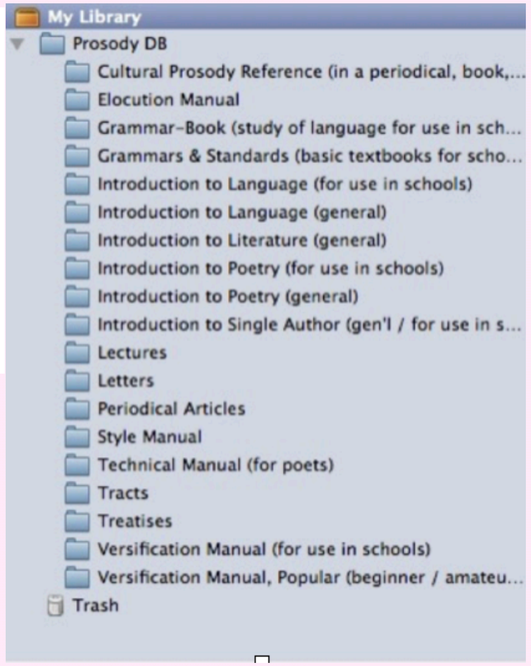
PRINCETON PROSODY ARCHIVE



ProQuest Start here.



Z

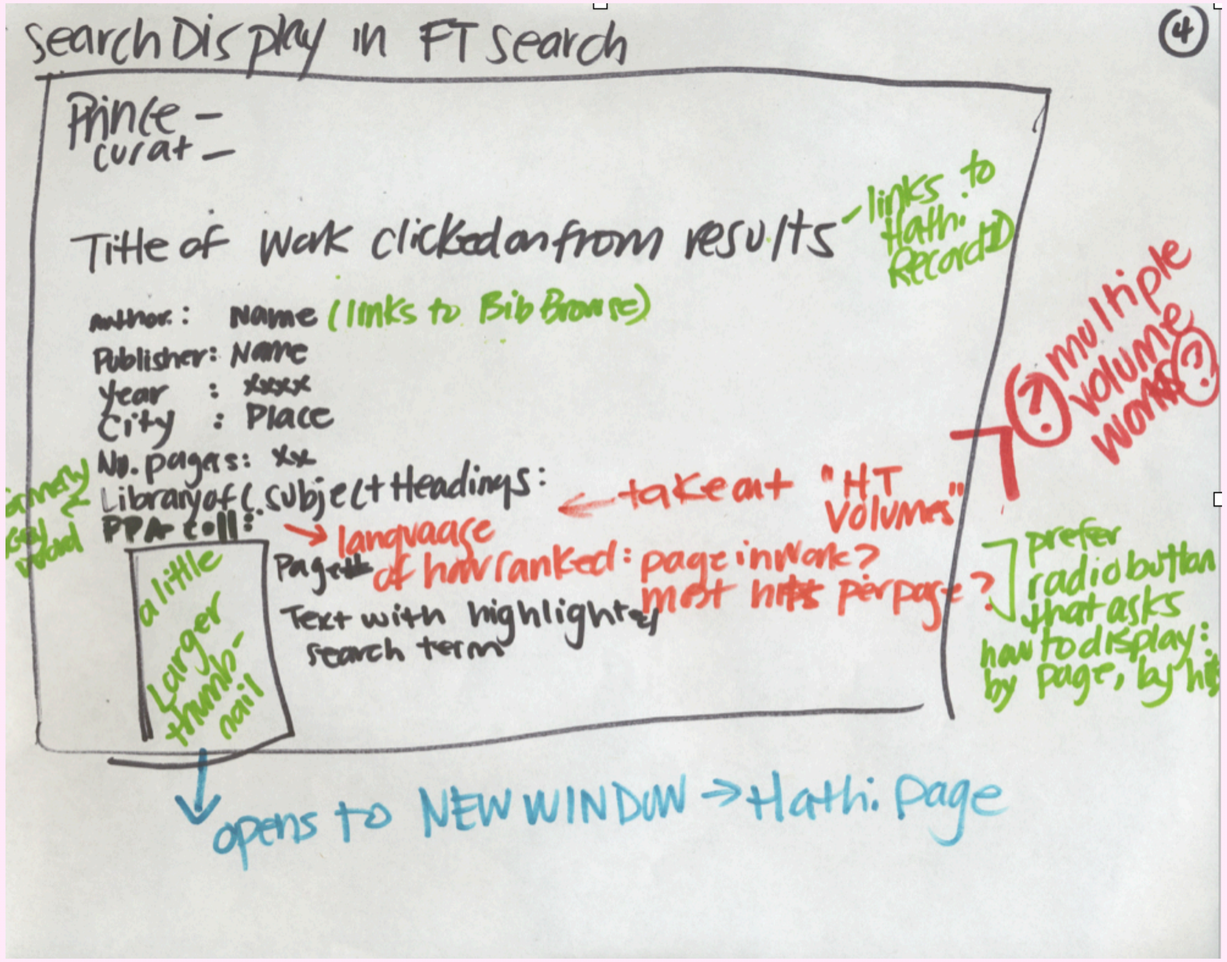


ProQuest Start here.

HATHI TRUST + GALE

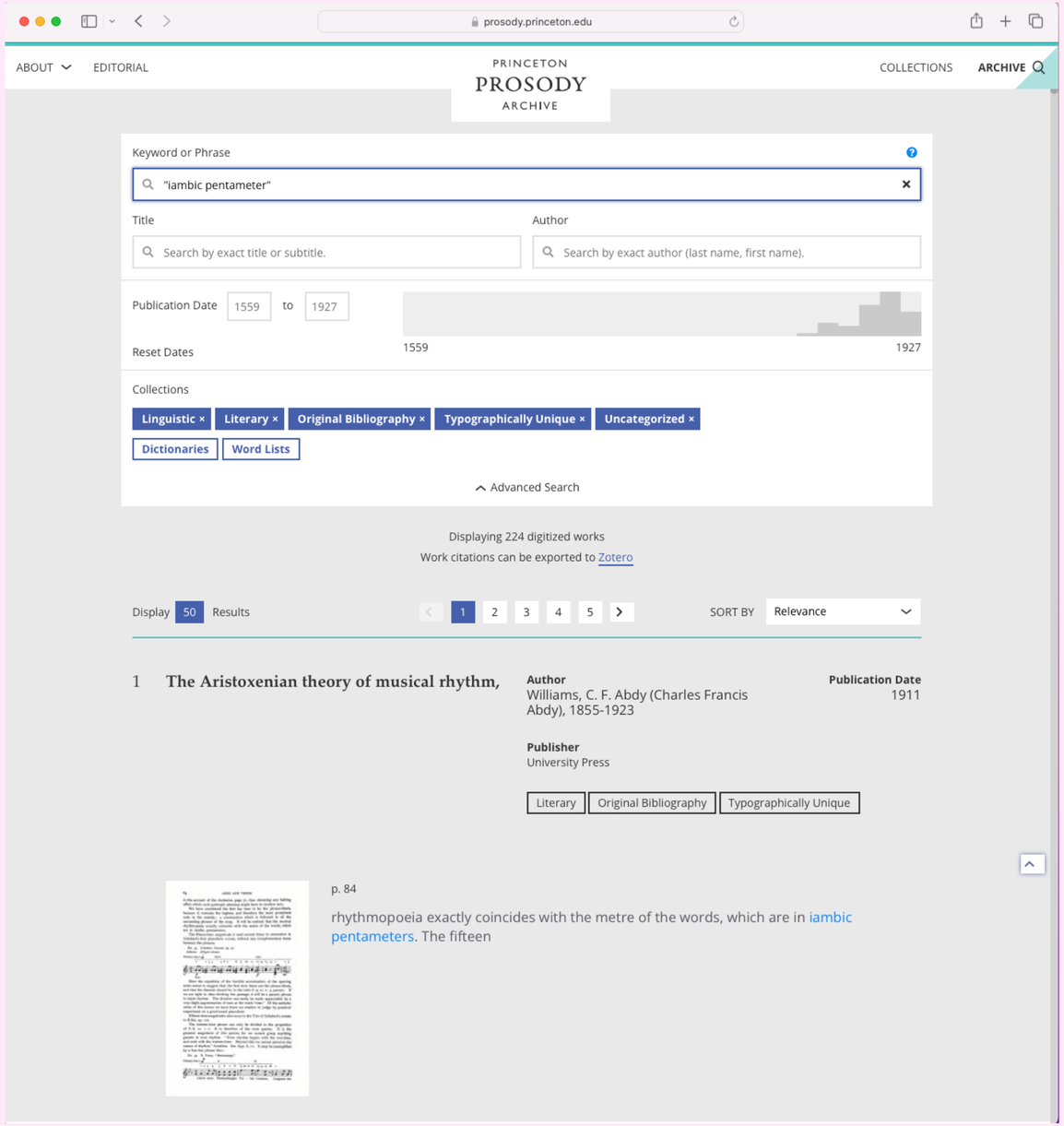
PRINCETON PROSODY ARCHIVE

How it started



prosody.princeton.edu

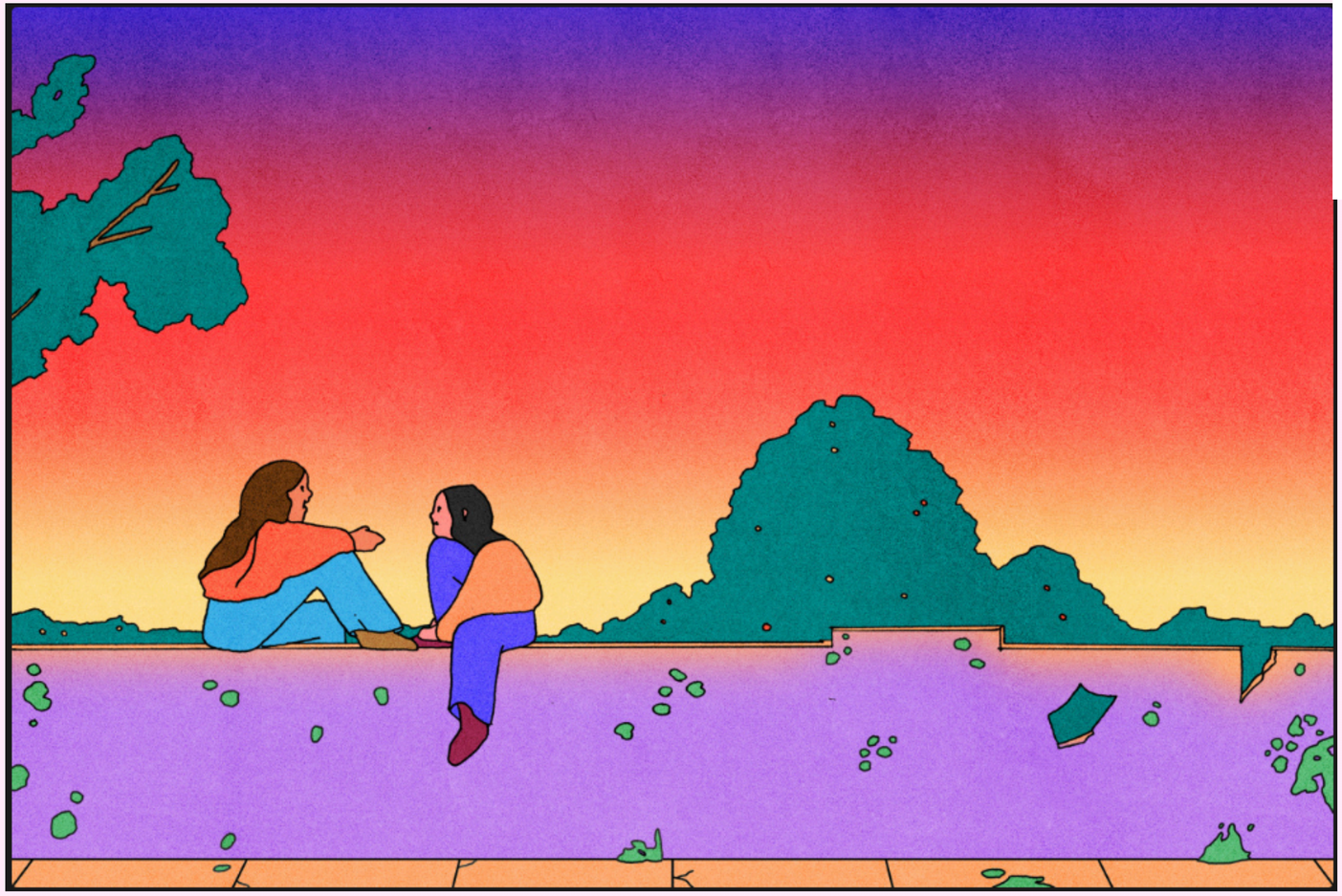
How it's going



Data Work allowed me to understand how much I had to learn about the history of poetry in relation to the broader history of linguistics and literary study.

Working with researchers at the Center for Digital Humanities showed me how challenging this work was for everyone: beginners and experts.

How do we come to an agreement about our data workflows if we don't have a culture in which we discuss and theorize our research practices — our assumptions about scholarship and research and writing — as humanists? How can we imagine honoring the interpretive labor of data work when people seem so worked up about even calling their source materials "data" to begin with?



Kevin 8:18 PM
I've been telling researchers to "develop a taxonomy" for categorizing data / answering research questions. Is "develop an ontology" more accurate?

Grant 8:19 PM
@Kevin not sure what the most accurate word is in terms of working with data, but "ontology" is definitely a more humanities-adjacent term than "taxonomy." (edited)

1

mm4 8:44 PM
I mean, we could all co-author it. It wouldn't be that terribly hard to gather what has & hasn't worked for humanities scholarship. We could get "short takes" or little practical example essays from a variety of people in different fields (David Mimno, ec)

2

Zoe LeBlanc 9:07 PM
So I'm 💯 all in on ALL of this @mm4! And also think more the merrier!! Should we set up a group brainstorm time??

rmunson 9:07 PM
calling you out on the fact that this NOT "random" - maybe should it go in a more durable channel or document?

1

rmunson 9:08 PM
mostly just want to make sure it's findable/searchable for when you return to it later

mm4 9:08 PM
STARTING GOOGLE DOC NOW TO COPY PASTE THANK YOU DOCUMENTATION QUEEN

2

rmunson 9:09 PM
I cannot suppress my love of documentation!!

budak 9:24 PM
where were you when the most significant monograph of 21st century DH was born in #random?

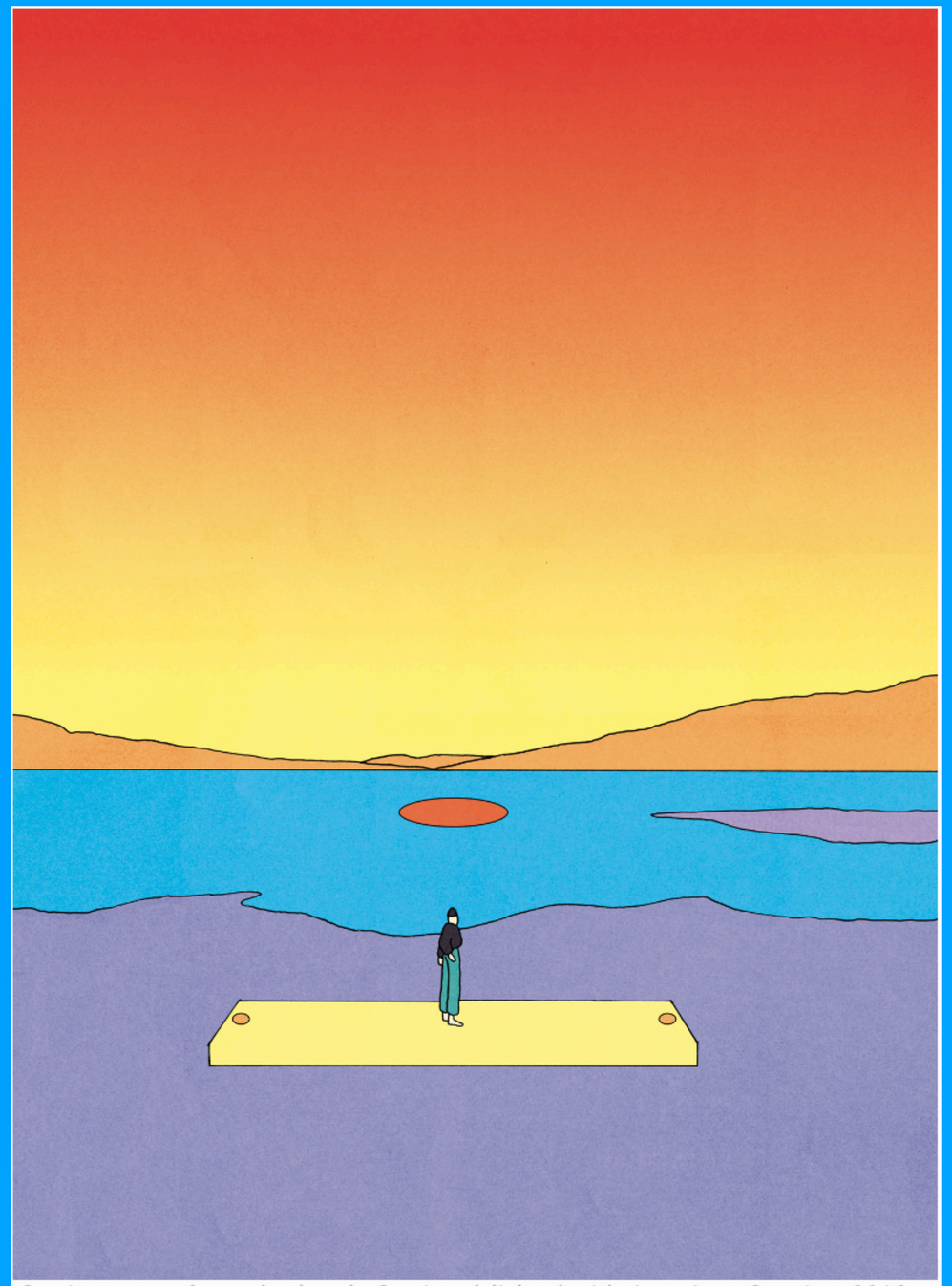
remember this moment

3

"There is no data without work. That is, there is no data without interpretation"

-Miguel Escobar Varela

Data Work aims to empower humanists and data scientists alike to dwell in the transformative possibilities of theorizing interpretation before tabulation -- not only what counts as data in the humanities or how we might treat material or non-material objects of humanistic study as data, but how we might bring more intention to our data practices through a deeper understanding of interpretation earlier in our process.

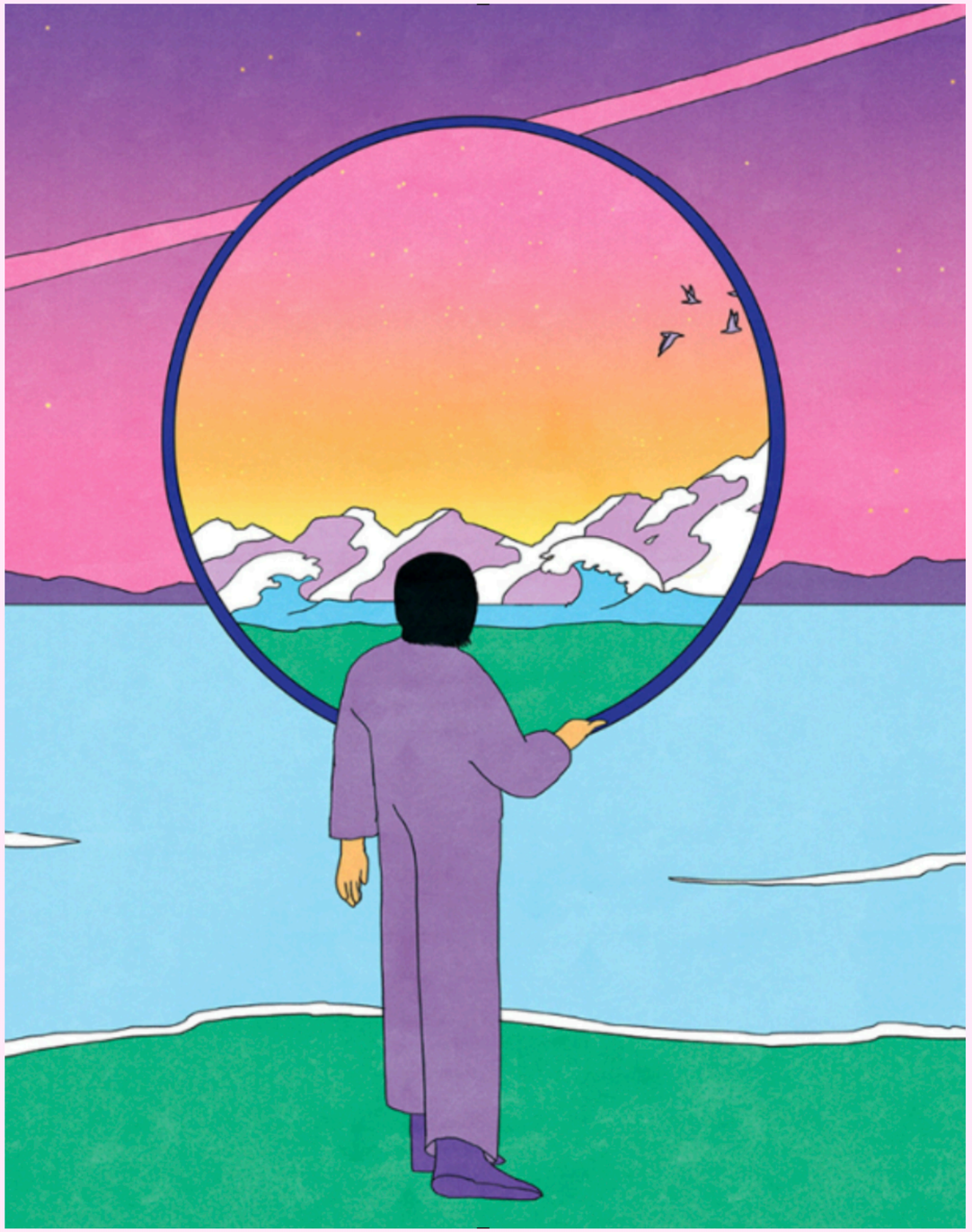


Cenit, extract from the book *Cenit* published with Apa Apa Comics, 2018

I am not going to define data we are past that we are all data humanists now, right?

For a source to be data, it must be made computationally tractable. By tractable, we mean following Willard McCarty, of humanist fame now, that information can be read, manipulated, and programmatically transformed by a computer or by computer-like operations. Computability is often accomplished by categorization. And these two qualities, categorization and computability, distinguish data sets proper from the sources that media scholars more commonly analyze and interpret.

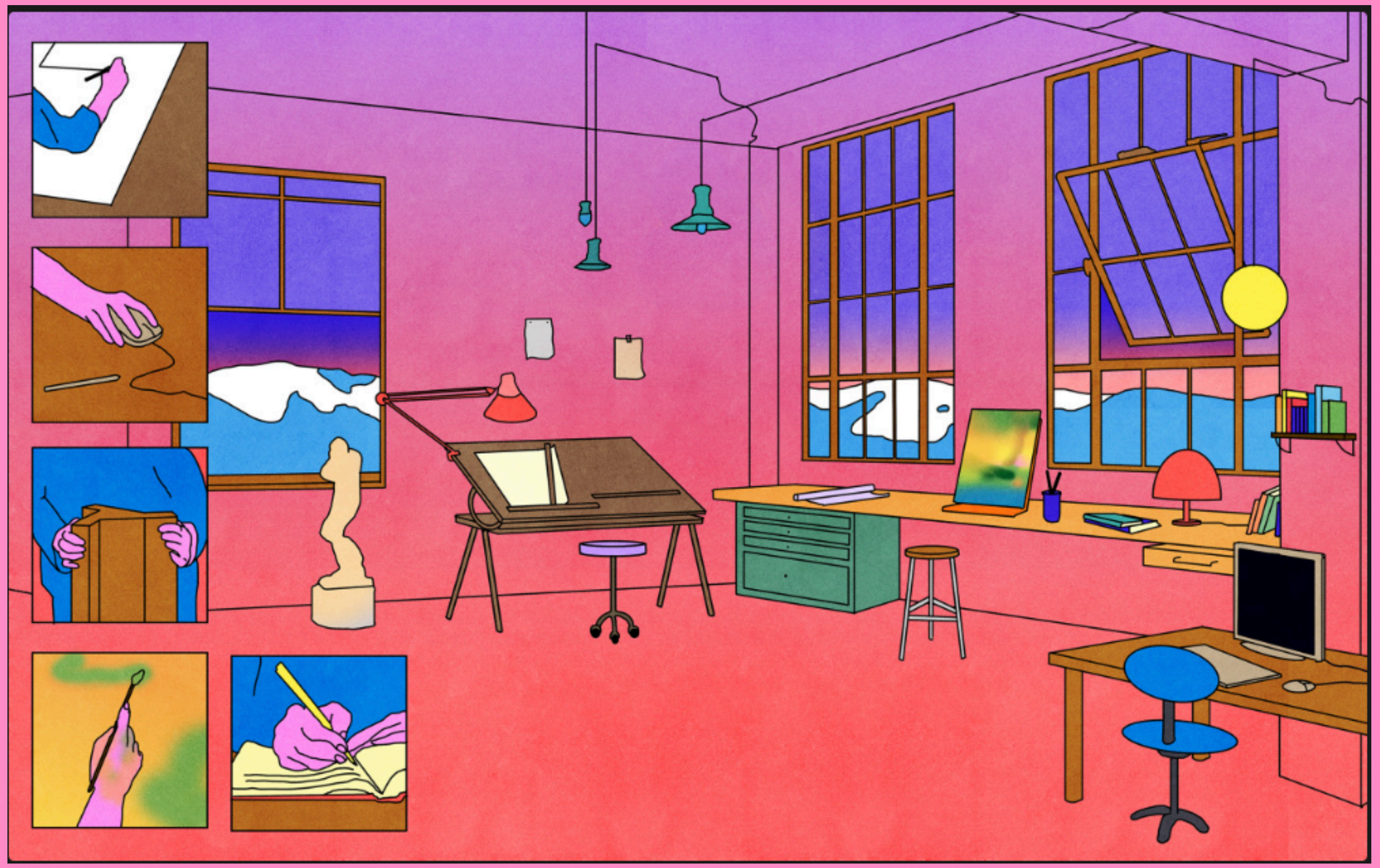
Lauren Klein and Miriam Posner “Editors Introduction: Data as Media” Feminist Media Histories 1 July 2017; 3 (3) 1-8

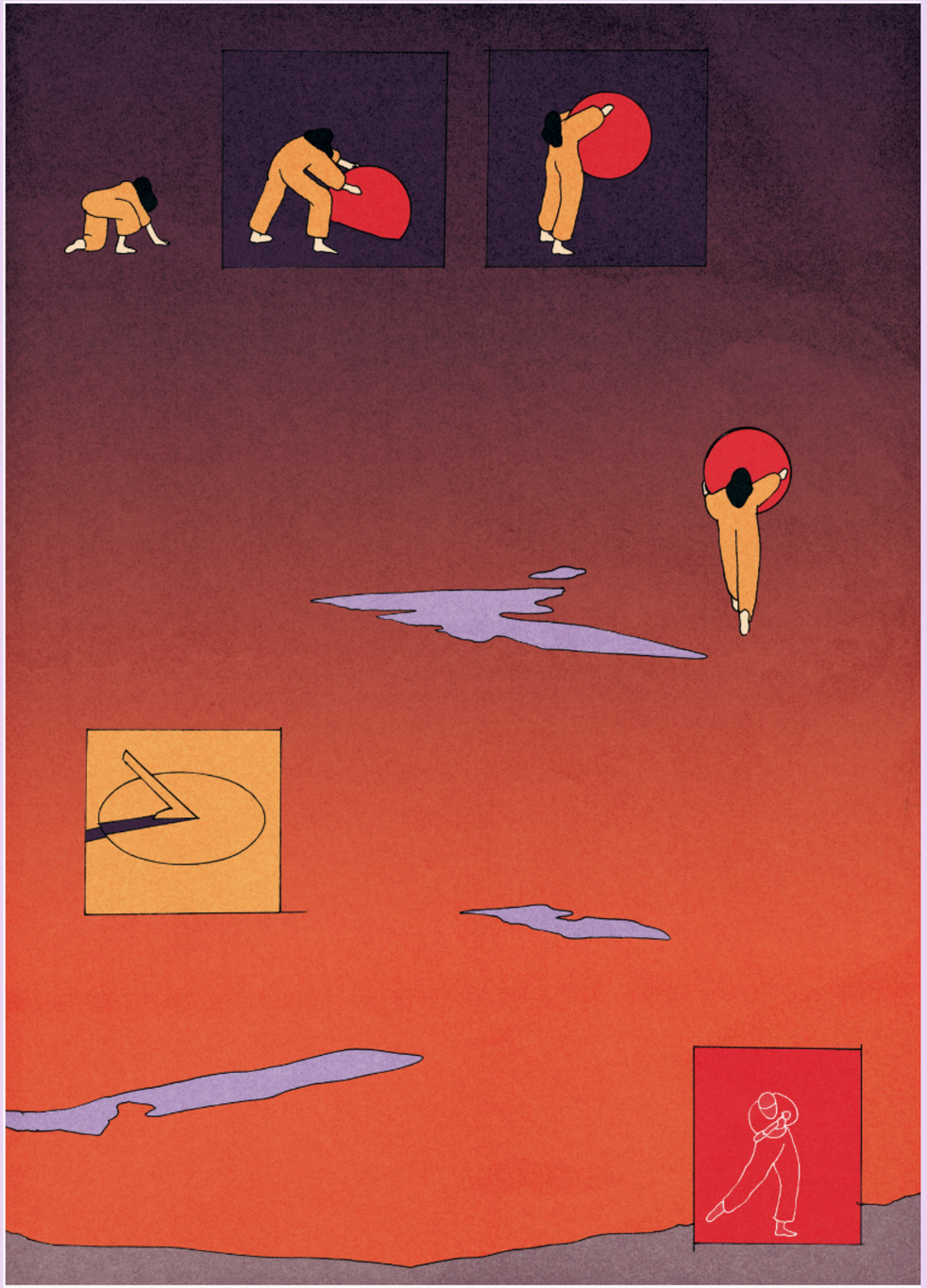


If I had to define data work as a narrower subset of the work that I do, it would have to describe all of the things that happen in the process of preparing data for use, right? So this includes identifying if you're starting from primary or archival sources, what those materials are, and not just which ones should I pick, but what makes sense for the questions I want to ask, then who has them and what form are they in? How much do they cost to get the rights, are they machine readable or not, all of this stuff, and then once you get them, what do they actually look like and are they useful for this type of analysis? Is there metadata that needs to be overlaid? How big are they and where do they go? Who needs to access them? I mean, there's a lot of stuff . . . data work is, at least in terms of the time . . . almost all of the work.

By the time you actually get to be doing the analysis part of things, you've usually done almost everything, and yet all of the things that I just described are not really clearly defined as things that matter or at least are counted, especially in the humanities.

-Lauren Klein, Emory University

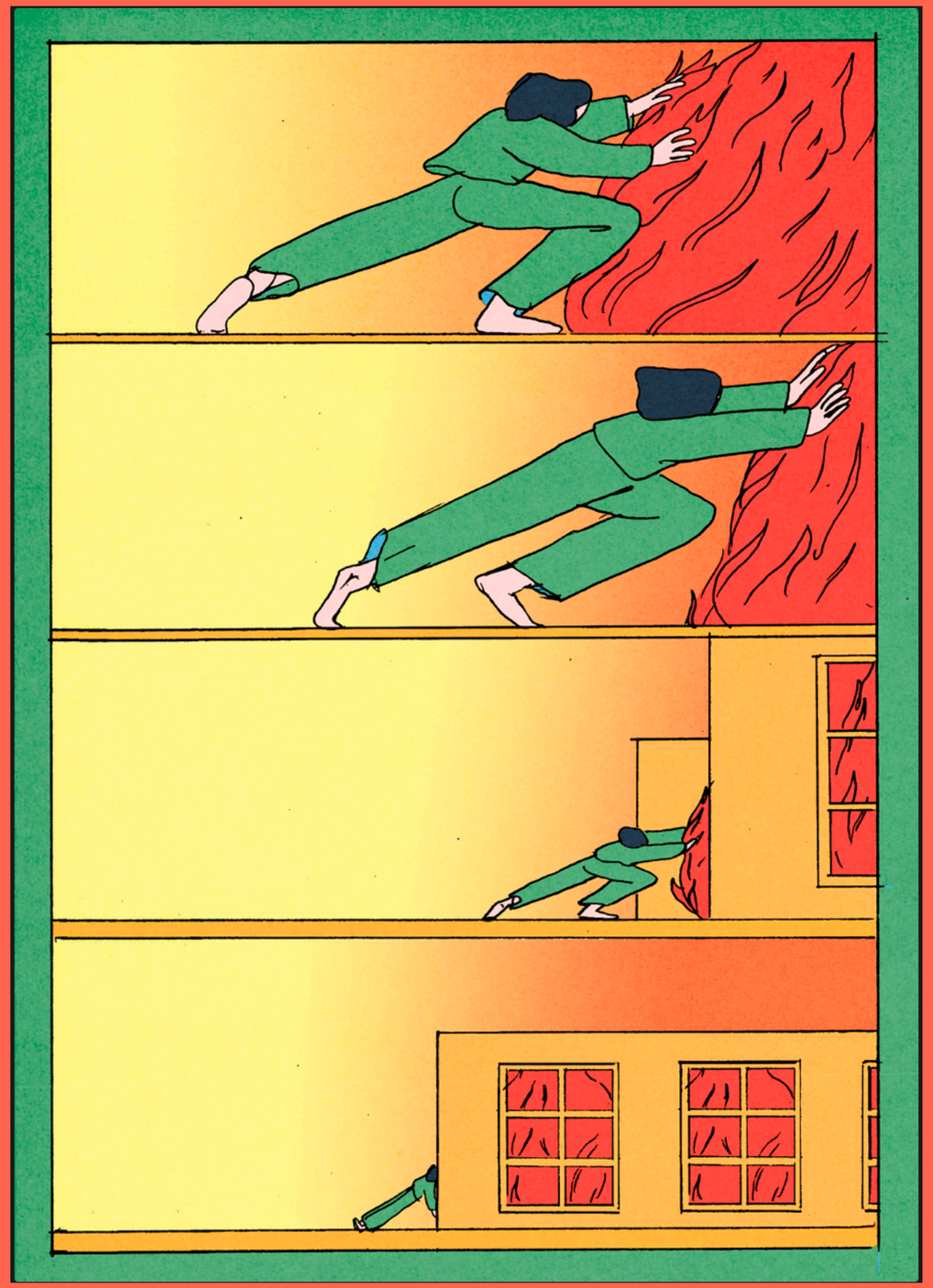




Exercise 08 for *The New York Times*

I think about the practice that I do in order to write. I think of it as a type of writing that goes on in order to do the writing that becomes the literary scholarship. I think I've thought about it in different ways over the time I've been doing it, so I think I started off thinking about it like cleaning, like putting things in order, and then I started thinking about it as curating, shaping and being conscious about the choices made, and now I'm increasingly thinking about it as writing in making, and making as well in a way that it's part of the material that has resistances and flows that make my arguments, that I have to work with them because they are part of what the texture of the argument is going to be.

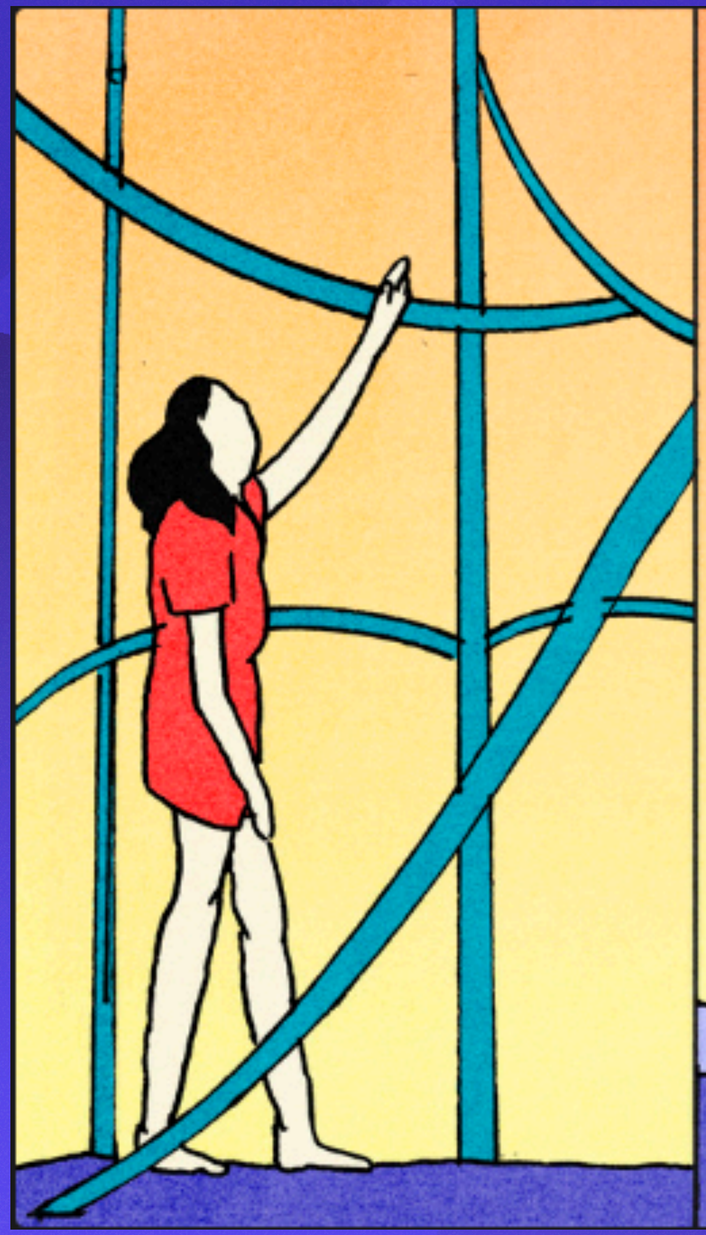
-Katherine Bode, Australia National University



One important piece of what data work means to me is I really like spreadsheets. I really do. I feel a little guilty about it because I think that they're really harmful in a lot of ways and I think that imagining that we can carve the world into entities that have attributes and that if one entity has a particular attribute, that that same category of attribute is appropriate for the next entity and the next . . .

And all of that is messy and gross and I don't want to be any part of it, but I also really do love spreadsheets. I do. And that tension between the gross mischaracterization of the beauty of the world and also the tremendously powerful clarifying effect of being able to talk with other people about entities and attributes and be getting to do what we can do when we are like, "I think that these nine things are entities and they're the kind of entity that I want to talk about in this conversation." It is so valuable. So the love that I have for what data work can clarify, it's in me and I have mixed feelings about it.

-Laurie Allen, Chief of the Digital Innovation Division (LC Labs) Library of Congress

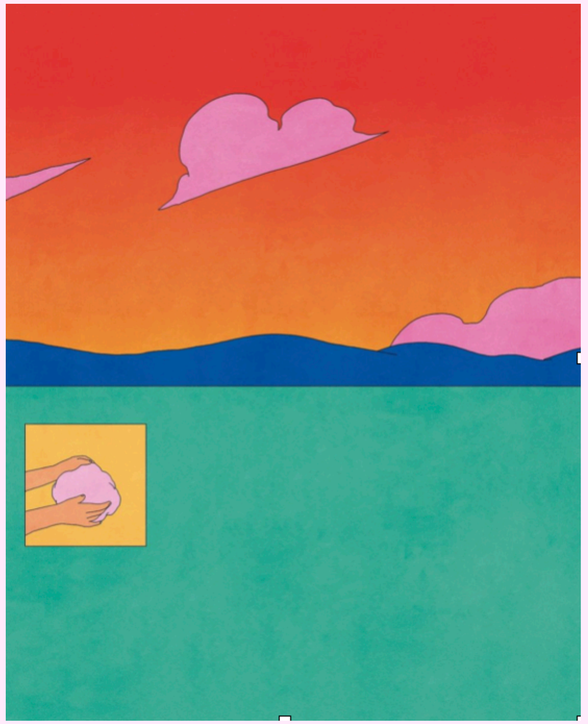


It's really useful intellectually to think of Data Work as a process where we're making choices about how we want to take this big, messy, complicated world and squeeze it into rectangular arrays of numbers so that it's machine actionable. And obviously that's not going to be right in some sense, but it's about saying very clearly what is it that is relevant to my current analysis, and what am I willing to let go of.

-David Mimno, Cornell

What is our relationship to this data and how are we making sure that these voices are centered in a way that they're able to hopefully speak what they wanted to say? And so for us, it is numbers, it is metadata, it is all of that. But it is also thinking about, if they had the opportunity to speak on their own, how would they do it? And are we honoring that? And then also our idea with data is making sure that it's clean. That it's clean in the sense that somebody else can take it and build something from that.

And we're always thinking about how a lot of these materials that have come to us, a lot of this data that comes to us or that we discover or we find or we encounter, it comes to us in a certain way that was, I don't know, preserved, collected, archived.



Our position is never to disparage and say, "Oh, well they didn't clean this or this wasn't done properly." It's . . . it got here. It got to us. Let's work from what was built. We try to figure out why it was put together in that way and then we try to build something from that.

-Gabriela Vaeza Ventura, University of Houston, Arte Publico

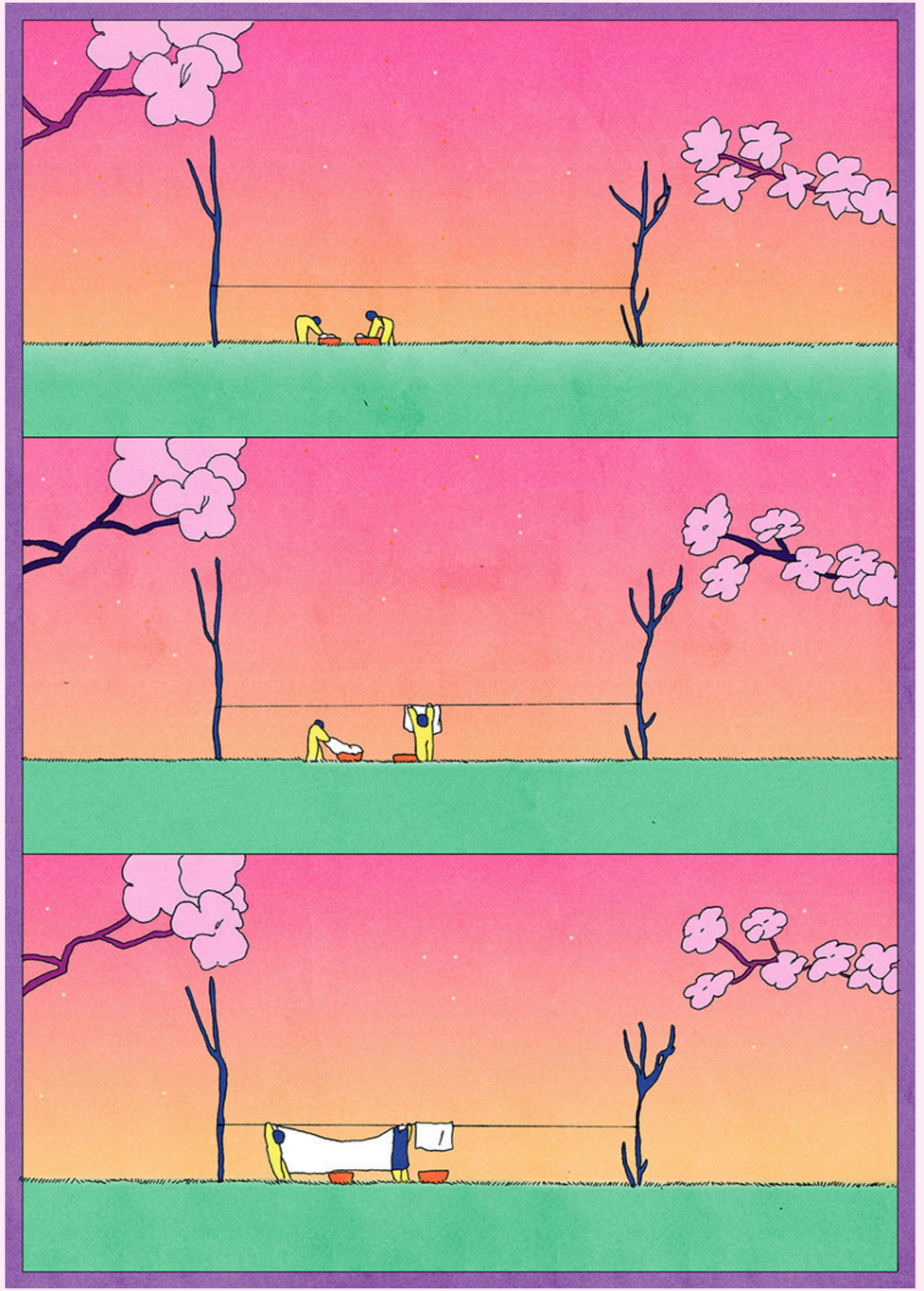
I’ve been actually wrestling with this a lot lately. Trying to find a sort, a non defensive way to take a position for humanities data. And it's difficult because there's always this sort of problem where when you're working to hopefully sort of expand the boundary of a question, you run the risk of just exploding or not even exploding, but denaturing the question to the point where there was no question. There is no distinction there or no categories, which sounds sexy at the outset but after a while isn't usually where we’re actually trying to head at any given moment. And so I say that because in thinking about data, I’ve been thinking a lot about what constitutes the data relationship between a person and how they survive.

Marisa Parham, University Maryland Collage Park,
African American Digital Humanities Initiative (AADHUM)

I think a lot of my data work is recuperating or even looking for data that has been missed or has been misrepresented, and looking for ways to represent it from the perspective of communities, bringing that subjectivity to what has been just a stereotype or has been created a single-story image.

And I think that's kind of a lot of my work in data, the data work that, how I interpret it, which is this process of looking for it, looking for what is missing and looking for different sources to bring those silences or that missing information, looking for ways to interpret it, to represent it, but also to take care of how, where is this data going to go from. Because of a lot of the surveillance, a lot of the vulnerability that it means working with communities from one side or the other, or with situations that mean telling names that cannot be mentioned, that cannot be public. I think a lot of my data work also involves being very careful about where is that data going to be placed, and how it's being categorized.

-Sylvia Fernandez Quintanilla, University of Texas San Antonio /
Borderlands, Archives, Cartographies



What does Data Work in the Humanities mean?

Considering the implications of our taxonomies & ontologies

Attention to the materiality & format of research materials, whether born digital or not.

Developing and maintaining relationships through collaborative & respectful, documented workflows

Care-work: creating relationships with communities who own the the data, creating and preserving access to untold histories.

Documenting decisions

Ongoing / Iterative

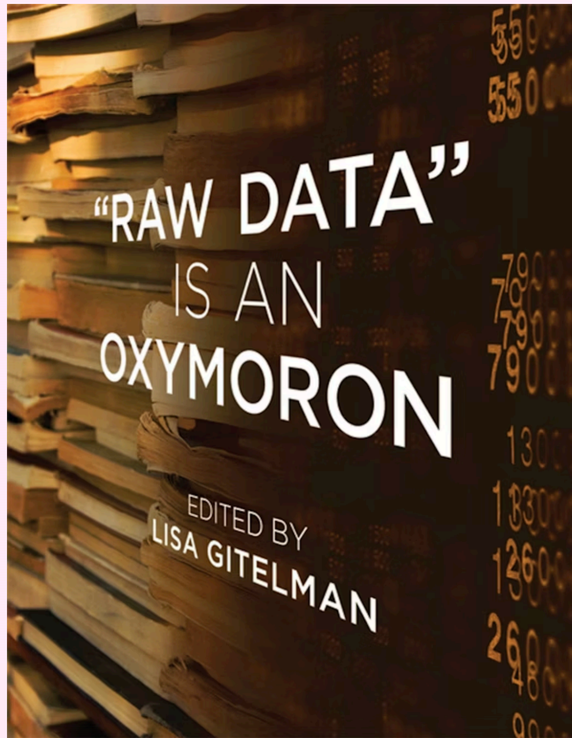
Understanding how data, technology, and data workers mediate access to our research materials

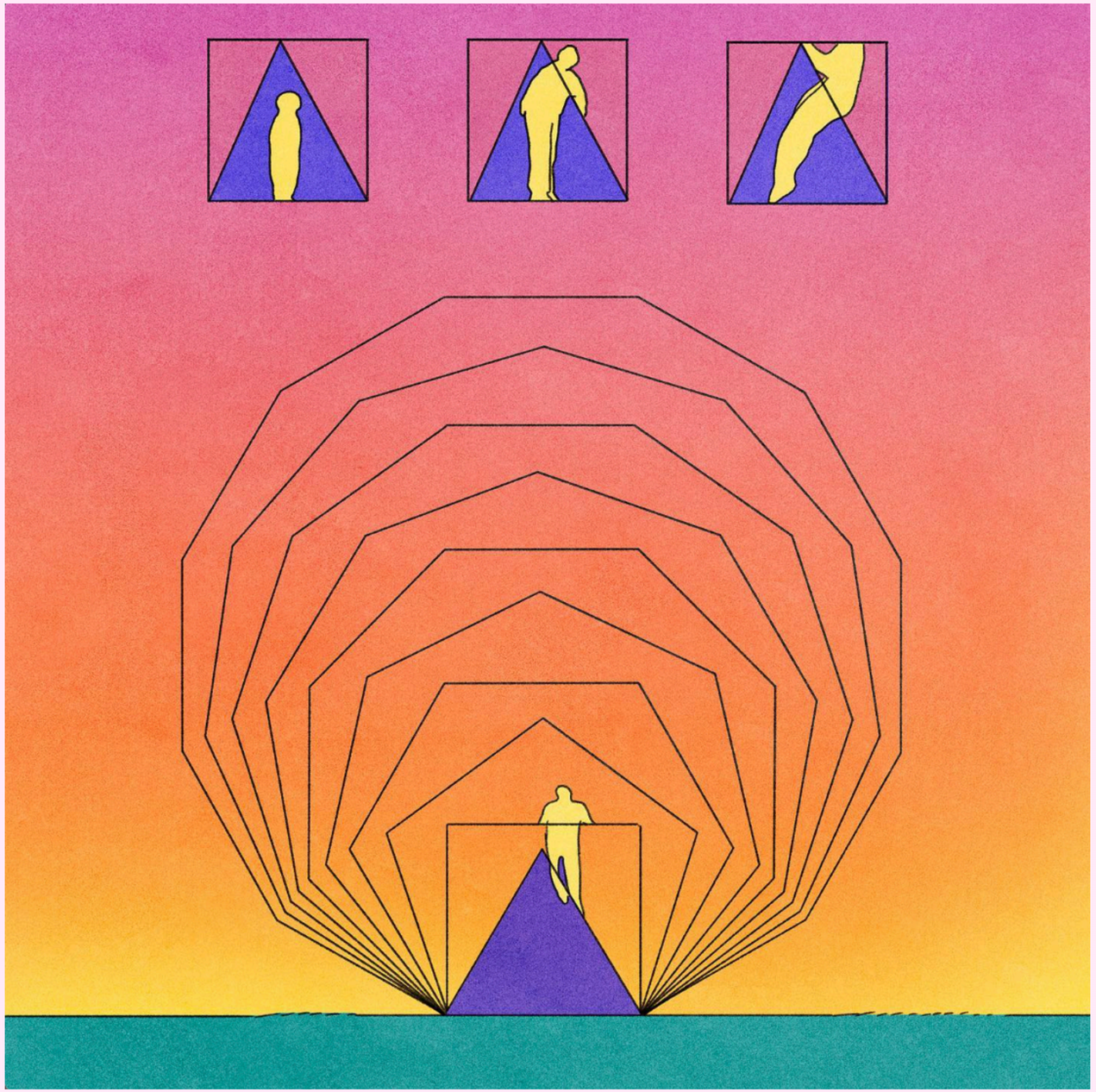
Labor intensive & labor conscious

Integral to Writing Process

Interpretive (but rarely acknowledged as interpretive)

Institutionally-situated / dependent





Thank you!