Digitize Again Forever

David Ribes

I am sitting in an office with shelves covered in physical media (see Figure 1). There are papers and folders and binders, CDs and tapes everywhere, stacked on top of each other, mixed and matched. Note the stamps, which are intended to label and categorize paper; and note the papers bursting from binders. Note the specialized drawers for CDs; and note the stacks of CDs on top. Despite the apparent mess this is not a disordered space, at least not for my interlocutor who knows, in broad stroke, which pile is which. She strides through the piles with confidence.

This is the data management center's main office for the Multicenter AIDS Co-hort Study (MACS). Since 1987 that data management center has been curating, preserving, cleaning, combining, and provisioning the data of the MACS. They are also responsible for digitizing the MACS's data again:

> We digitized these data in the early 2000s. You can see the CDs behind you [she gestures to shelves behind me], which we have to *digitize again* soon because CDs only have a shelf life of about 10 years. They are overdue. (emphasis added)[1]

I've sat in other rooms like this before. Since the MACS has three other geographic sites across the US, there are also three other rooms that look somewhat akin to this one. The piles, the heterogenous media, the definitive sense of order which is nevertheless inscrutable to my eyes.

This room is in Baltimore, Maryland, USA, but this detail doesn't matter very much to the tales I will tell here. I find that the rooms of information managers across my different research projects all look much the same. Even when it's not the MACS, and even when it's not about AIDS, I've still been in rooms that look much like this.

---

1   Anonymized interview with MACS Information Manager, October 2011.

Figure 1: Center for Analysis and Management of MACS Data

Source: Photograph by the author, all rights retained.

The MACS was founded in 1984. Then, data were first inscribed on paper forms, quickly transferred to long magnetic tapes and (for them) startlingly compact floppy disks. Those data on paper or disk were shared across its four geographic sites by fax and by mail, but only after first phoning someone to confirm that the data were 'coming by fax' or 'in the mail'. That was all a long time ago and a lot of things have changed since, such as the disease they study (fatal in 1984, a chronic parasite today), their scientific instrumentation (which regardless, almost all end up outputting data), and of course all the media that they use to store these data. Nothing has stayed the same, except of course that the MACS keeps producing more data, and then digitizing it, and then digitizing it again.

After the $5\frac{1}{4}$ inch floppy disks there were the even more compact $3\frac{1}{2}$ inch disks, hard drives (which were also placed in the mail for a time), slightly larger but somehow more compact discs, and of course The Internet. Of course, The Internet did not pass these information managers by, they kept up. But despite its tidy name The Internet is not one thing, and it has done nothing to end the parade of digitizing their data again. There have been shared networked drives; carefully guarded servers run by departments, colleges or the university; even more carefully guarded private provider servers; and of course a recent bevy of Cloud services. All of this has been The Internet, and none of it has slowed the MACS' efforts at digitizing their data again:

The last time a vendor told me they were going to "digitize our data" [curling her fingers around the words], I guffawed out loud. I know it was rude, I couldn't help it, I'd heard it a few times before.... [chuckles] Anyways, we did end up hiring them, and they did digitize our data again.[2]

How Digital Can We Get?

Here I reject what I call the saltationist model of digitization (James 1909; Latour 1999). My argument is that being digital is itself not a binary state, a singular jump from one condition to another, despite how much the digital itself insists that is the case. Instead, I consider digitization to be hungry, bottomless, always looking for its next meal, forever dissatisfied with everything in the past, including its own past efforts of digitization. The children of the digital, those dubbed 'born digital', are not exempt, they are just as tasty, perhaps more so, and just as subject to digitization's endless maw. Like Saturn, digitization will consume them too, and then once again.

I have thought that digitization is an endless hungry maw for some time now. I have been studying scientists digitizing their data again since about the early 2000s. As we saw above, in the past I've worked with biomedical scientists and their data; I've also worked with geologists, ecologists and physicists digitizing their data again, amongst others.[3] I should know by now that digitization is not a binary state, more than for any other reason because sometimes I have returned a decade later to the very same group of scientists only to find that they are once again digitizing their data.

Often, they don't call it digitizing their data, there are other words. Such as 'implementing' a metadata standard, or 'registering' data to a computational ontology, or 'harmonizing' across heterogeneous data. In this I am not fooled though, I am theoretically sensitized (Glaser 1978) enough to recognize a common pattern when, regardless of what they call it, really they are just digitizing their data again. In this, the central theoretical term of this edited volume, "datafication," too smells to me suspiciously like digitization again.

But even with all this critical distance, I still find myself regularly duped by the promissory pitch: 'just try it once more' digitization coos in my ear, 'this time will certainly be The One'. The promise of digitization is that, going forward, things will never be the same again. For digitization, The One is the occasion on which being digital has been achieved, forever. This is how we often speak of digitization, a singular transformative leap from some anterior condition (such as "ana-

---

2   Anonymized interview with MACS Information Manager, October 2011.
3   Ribes and Polk 2015; Ribes and Bowker 2009; Ribes et al. 2012.

log," sometimes "material," or perhaps "qualitative") to a new state of digital being. It is a one-way trip to becoming binary and symbolic, letters of light, electricity, and magnetism. Data.

I know this claim cannot be true. I argue against The One transformation in many of my papers, (including this one, such as with the bit of ethnographic evidence above). And yet still, I am not immune to the siren call of digitization. I often head straight towards it, convinced.

A small example via a true personal story: Just recently, about 10 years ago now, I digitized my music collection. My collection was mostly CDs. As these discs appeared to be on their way out the door I adopted a portable MP3 player, and then soon, a smartphone to replace that too. As a practice, digitizing my music collection looked like this: I would insert the CD into the drive and using software ("ripper") converted the CD's native format (WAV) into the more compact, but slightly lossy, MP3. Then, I repeated this task with the next CD. It took weeks.

Obviously, the CDs were "already digital," WAV files are data (Sterne 2012). Even still – and equivalent to the data CDs of the MACS above – the software I used called it digitization and so too did the guides I looked up online for how best to "Digitize your music collection." Technically, something digital cannot be digitized, in practice we digitize again all the time.

Despite all the skepticism I have communicated above, I really did digitize my music thinking that this time was The One. That, after all the musical media transformations I had gone through (...vinyl, tape, CD...), now I would no longer need to transition again because now my music would be digital. Files of light and magnetism that I could take, no matter what, with me to wherever the digital went next. Obviously, I was wrong; but it's *how* I was wrong that I think is of interest. Digitization proved surprisingly discontinuous from its own past.

In my defense, I was not wholly naive. I did bring to bear some of my academic craft to the task. Such as thinking to myself that certainly the future would still involve a great deal of maintenance and repair, a practical labor of transition. Perhaps MP3s would go out of date, and I would have to spend some time converting (i.e., digitizing again) those files to the new format. But I still thought that the files themselves were now eternal, just a matter of converting on to the next thing. In this I was wrong; this is not at all what happened.

That digitized music collection – stored as hand-crafted MP3 files – today sits saved on a hard drive that itself sits on top of the original CD collection, in my garage, in a box, untouched for nearly a decade (see Figure 2). My prediction above – that this digitization was finally The One, or that I would simply convert these files to the next format – did not pan out as such. I did not change these digital files into anything because the next form for digitization turned out to be a completely different kind of digitization, wholly exceeding my predictions as is often the case with the wily ways of digitizing again.

Figure 2: The author's digital CD music collection, digitized again onto a hard drive, which resides in this box, in the garage, untouched for a decade, until he took this picture.

Today my music collection is provided to me through The Cloud. I pay a monthly fee to have access to the music I listen to. I never digitized my collection, at least not in the sense of 'uploading' my MP3s to The Cloud or some other form of direct converted continuity for these files. Instead, in this case, digitizing my music again involved selecting albums and tracks from a vast drop down list of "almost all music." It took weeks.

I'd be comfortable saying that this new collection is equivalent to my old collection. But it has also grown, shrank, and some tracks just don't sound the same. The Cloud, as with CDs and MP3s, and all other digitizations that came before, insists that this is the end-state for my music collection, forever.

The Saltationist Model of Digitization

A saltation is a jump, perhaps a particularly vigorous jump. In the sciences saltation is used metaphorically to indicate a dramatic, singular transformation, such as in biology where a saltation is an abrupt evolutionary change. But while there *may* be saltations in evolution, digitization does not occur as a singular decisive and transformative leap. Instead I consider digitization to be a bottomless discontinuous chain, and I name my oppositional foil, the saltationist model of digitization.

I draw this term from actor-network theorist Bruno Latour, who in turn borrows it from Pragmatist philosopher William James (James 1909; Latour 1999). Latour and James too use "saltation" to denote their foil: a vast set of epistemological and ontological commitments that we might call representationalism or objectivism (Ribes and Sutherland 2021); a position that asserts a world "out there," its representation "here," and the gap between these two as the central matter for concern. Saltation in Latour and James broadly evokes classical modernist philosophical distinctions between substance and form, matter and mind, and most notably world and representation. I will not tackle these vast theoretical commitments in this little essay, but I too intend my use of saltation to evoke how digitization has often been miscast in the idiom of representationalism.

Here, I use saltation more modestly to define the model that casts digitization as an historical binary, a once-this-then-that, single-instance-transformation. Instead I argue that *digitization is, in principle, a bottomless chain of discontinuous equivalencies centrally concerned with continuing to exist.*

I use the term *equivalency* in the sense imparted to that term by actor-network theory: one thing taken to be the same as another. More strongly than 'taken to be', that if something or someone were to dissent (Latour 1987) as to that similarity, the equivalency would fight back. Equivalencies are real, in the Pragmatist sense of that which resists. When digitizing the data of the MACS from magnetic tape, to disk, to disc, to The Internet, it is always the same, equivalent. If MACS scientists did not think equivalency had been achieved, then it would not be digitization at all, merely dangerous and disposable garbage data. If I, or anyone else, was to suggest that those data are not the same, the MACS would fight back. The equivalencies of the before-and-after of digitization are very real, but each era of digitization is discontinuous from the last.

I use the term *discontinuous* in the sense that digitization-next regularly proves to be distinct from digitization-before. Digitization itself denies this, it will tell you that it is always just ones and zeros, which is what it was before and what it will be next. But approached as practice – that is, respecified (Garfinkel 1991) – digitization-next is rarely like digitization-before.

Recall that when I first digitized my music collection (ripping) it turned out to be quite different than the second time (selecting music from The Cloud). So too

for scientists and all their legacy data. This next tale is no different in its form, even though it is more technical (by which I mean, less easy than ripping): Once I observed geologists over months as they sought to digitize their data again, on this occasion by "registering" their data to a "computational ontology." At the beginning of this process they already had some of the best organized data out there: in a relational database, with plenty of descriptive metadata, readily available online via a Web service. Even still, something was lacking, such as being able to conduct a semantic search, something ontologies would allow them to do. But registering to an ontology demanded something additional, something more, i.e., formally encoding their geoscientific knowledge into the predicate logics of the Ontology Web Language (OWL), and then registering their data, column by column, to that domain-specific ontology. In the paper I collaboratively wrote from watching these geologists, we argued that geoscientists had to "learn" (Ribes and Bowker 2009) to articulate their informal knowledge in the predicate logic, because even though these geologists were already highly data literate, and their data were highly organized, open access and online, in the face of these computational ontologies they still had to learn how to digitize their data again, a discontinuous practice from what they had done before.

Lastly, I use the term *bottomless* somewhat colloquially, like a Bottomless Mimosa: in principle there is always more to be had, in practice there are concrete limits; which is also how I mean *forever*. Digitization is bottomless in that there will always be another instance of digitization just around the corner, demanding, needed. The next form of digitization will always insist on something new, something more. Data will need to be relational or maybe object-oriented; there will be a need for more or better metadata; they must be uploaded, online, registered, open access, reproducible, differentially privatized… The point is that sooner than later data are 'overdue' and at risk of being lost. Respecified as practice, the next digitization will always be different than what came before, even while digitization will always insist it's just a matter of ones and zeros, a one-off transformation like before again.

Clearly, there is some element of a corporatist motive here – as with the vendors for the MACS who forever promise a final digitization only to return a few years later with a new one; or, as with The Cloud provider that I now pay monthly for my music. There is money to be made by promising The One transcendence even while digitizing again, forever. But there is more to digitization's bottomless hunger than greed; it is also about continuing to exist.

In principle the data of the MACS, stored on CDs, or my music collection, also stored on CDs, could continue on indefinitely, with the sort of maintenance and repair I suggested above: tweaks along the way. But in practice I think these data, left on CDs, would cease to exist, at least practically. Consider, if today an HIV scientist requested data from the MACS and was told that those data would arrive

by mail, on a CD, in a comma-separated flat-file, in a couple of weeks; I think that researcher would at least consider looking elsewhere for data, such as via The Internet. Even if 20 years ago they would have regularly waited two weeks for comma-separated data, today many would not. Personally, I have become used to waiting no more than a second for my music, even if I know that the track that plays from The Cloud may not quite be the one that would have played from the CD in my collection. In principle, well-kept data are eternal, in practice they must be digitized again, not because the data have changed, precisely, but because they cannot stay the same while everything else does not.

To say that digitization is a bottomless chain of discontinuous equivalencies which are centrally concerned with continuing to exist, is a long way of saying digitization is a parasite – in the sense imparted to that term by poet-philosopher Michel Serres. In order to continue to exist, and despite all promises of a one-time transcendence, digital things, such as data, but also everything and anything else, must be interrupted, transformed and renewed, now, and then again later, forever:

> a new obscurity accumulates in unexpected locations, spots that had tended toward clarity; we want to dislodge it but can only do so at ever-increasing prices and at the price of a new obscurity, blacker yet, with a deeper, darker shadow. Chase the parasite – he comes galloping back, accompanied, just like the demons of an exorcism, with a thousand like him, but more ferocious, hungrier, all bellowing, roaring, clamoring. (Serres 2013)

The Parasite can be read as an essay on information – "In the beginning was the noise," Serres writes – but it can also just as easily be read as about datafication, The Internet, The Cloud, or digitizing. They are not that different. More accurately The Parasite is about information again, datafication again, or uploading again, which are all equivalent discontinuities for the same thing: digitize again forever, "the chain seems unending" (Serres 2013).

References

Garfinkel, Harold. 1991. "Respecification: Evidence for Locally Produced, Naturally Accountable Phenomena of Order, Logic, Reason, Meaning, Method, etc. in and as of the Essential Haecceity of Immortal Ordinary Society (I) – An Announcement of Studies." In *Ethnomethodology and the Human Sciences*, edited by Graham Button, 10–19. New York: Cambridge University Press.

Glaser, Barney. 1978. *Theoretical sensitivity*. Oakland, CA: University of California Press.

James, William. 1909. *Pragmatism and four essays from the meaning of truth*. Cambridge, MA: World Publishing Company.

Latour, Bruno. 1987. Science in Action: How to Follow Scientists and Engineers Through Society. Cambridge, MA: Harvard University Press.

Latour, Bruno. 1999. *Pandora's Hope : Essays on the Reality of Science Studies*. Cambridge, MA: Harvard University Press.

Ribes, David, and Geoffrey C. Bowker. 2009. "Between Meaning and Machine: Learning to Represent the Knowledge of Communities." *Information and Organization* 19 (4): 199–217.

Ribes, David, Steven J. Jackson, R. Stuart Geiger, Matthew Burton, and Thomas Finholt. 2012. "Artifacts that Organize: Delegation in the Distributed Organization." *Information and Organization* 23 (1): 1–14.

Ribes, David, and Jessica Beth Polk. 2015. "Organizing for Ontological Change: The Kernel of an AIDS Research Infrastructure." *Social Studies of Science* 45 (2): 214–241. Accessed September 27, 2021. https://doi.org/10.1177%2F0306312714558136.

Ribes, David, and William Sutherland. 2021. "Representationalism and Digital Imagery." In *Postphenomenology and Imaging: How to Read Technology*, edited by Samantha J. Fried and Robert Rosenberger, 283–292. Lanham/Boulder/New York/London: Lexington Books.

Serres, Michel. 2013. *The Parasite*. Minneapolis, MN: University of Minnesota Press.

Sterne, Jonathan. 2012. *MP3: The meaning of a format*. Durham, NC: Duke University Press.