From: Narrative Intelligence Lab
Subject: Memo 003: What is Digital Ethnography?
Date: December 29, 2025

CUØ /'kju:nɪl/

Setting aside the long and over-theorized history of ethnography as method, this note advances a deliberately partial and pragmatic account of doing ethnographic work online. Our approach combines qualitative and quantitative techniques while extending observational fieldwork beyond interaction and representation to include media, software, and platform analysis as first-order objects of inquiry.

Ethnography

To think clearly about ethnography online, it helps to be explicit about what ethnographic fieldwork ordinarily demands. For our purposes, ethnography organizes itself around at least three complementary imperatives under the umbrella of fieldwork: to observe, describe, and participate.

These do not unfold sequentially. They operate as co-present demands, each shaping how the others proceed. Observation calls for attention, curiosity, patience, and an eye for detail. Description introduces judgment—questions of interpretation, selection, and even literary form. No matter how thick the description, it cannot capture the fullness of experience. An ethnographer's gaze remains bounded by position and point of view. And yet, especially in the early stages of fieldwork, I often do not know what I am looking for. Details that later appear incidental or superfluous often enter the record without their significance yet being clear (Malinowski 1922; Emerson, Fretz, and Shaw 1995).

Description thus both exceeds and falls short of experience. It overwrites and underwrites at the same time, unfolding within a tension between what comes into view and what remains illegible, between what I notice and what only later coheres. Moving through familiar territory, smoothed and rendered invisible by repetition, I place myself deliberately in the position of a visitor—pretending to see things again for the first time, restoring a sense of friction and detail necessary for later analysis.

Never definitive, ethnography records a singular encounter, awash in collective experience. This balancing act between subject and object demands above all participation. The observer cannot detach cleanly from what comes into view. To learn something about others requires stepping outside oneself and putting oneself on the line. Doing so entails becoming exposed, complicit, and vulnerable to experience. The line between subject and object blurs in the process. Attempting to resolve that tension, leaning too far to either side of it, lessens the project. Complete identification courts hubris; claims of scientific detachment collapse under scrutiny (Clifford 1986; Marcus 1995). Ethnographic observation rather proceeds in the tension between these positions, sustained rather than resolved.

Ethnography Online

Digital environments do not eliminate the imperatives of fieldwork; they redistribute them. Crucially, ethnographic practice no longer coheres around a shared material setting. The observed surface conditions online increasingly fall beyond reach, into media systems, software architectures, and platform infrastructures. Surface appearances may differ significantly from the conditions that make them possible, particularly under conditions of unequal access, platform opacity, and methodological constraint (Hargittai 2020).

Online, observation rarely coincides with co-presence. Interfaces format what comes into view, platforms filter it, and computational systems constrain it in ways that remain largely inaccessible to both participants and observers. Participation likewise unfolds through personal profiles, reputational systems, algorithmic curation of content, and terms of service, often detached from the bodies that generate them and governed by rules enforced in code rather than articulated in practice. Description, in turn, confronts a field in which much of what structures social life operates below the threshold of perception. To describe online social life critically requires moving fieldwork beyond screen appearances to include the media, software, and platforms that actively delimit the possibilities of action, interaction, and meaning.

In physical settings, the material properties of a specific place delimit the scope of ethnographic description. Consider a stadium sport. A football match unfolds within visible and shared constraints: the field, bodies in motion, and the published rules of the game. When that same game is played digitally, those constraints no longer coincide in the same way.

The digital field persists as an image on screen, while its material instantiation—hardware, storage, and networked infrastructure—recedes from view. The body appears as a controllable avatar, governed by the physics of the game engine rather than the capacities of the player's own body. What the avatar can do follows one set of constraints; what the player's body can do follows another, shaped by interfaces, ergonomics, latency, and physical circumstance. Though the rules of the game remain legible, they now operate twice: once as an explicit system of play, and again as software code that enforces, supplements, and occasionally overrides them in ways opaque to both players and observers.

In digital play, what appears unified in action no longer arrives as a single analytic object. Representation and constraint diverge, and the material conditions that organize action recede from immediate observation. The difficulty posed by this example is not that digital play lacks materiality, but that its material conditions no longer arrive together with their representations. The apparent unity of action must therefore be critically decoupled or delaminated if it is to be described at all. The task of digital ethnography is not simply to notice this separation, but to develop methods capable of tracing it.

At the level of representation, the digital football match presents itself at the surface: the camera angle, the movements of avatars, the scoreboard, the timing of actions and outcomes. Here, ethnographic practice remains closest to its familiar form. Description, close observation, and comparison still apply, drawing on long-standing traditions that treat representation as both a site of meaning and a problem of mediation (Geertz 1973; Emerson, Fretz, and Shaw 1995; Clifford 1986).

At the same time, representation in digital environments arrives already structured as data (Shaw and Hill 2014, 2018). This condition invites the augmentation of description through computational analysis, not as a replacement for ethnographic judgment but as a means of extending it across scale and repetition (Yi Tenen 2017).

Screen captures, replay files, logs, and interaction traces can be collected and compared systematically, allowing patterns to surface that remain invisible in isolated observation. In practice, this may involve recording multiple matches, annotating moments of play, and comparing how similar situations unfold across players, teams, or difficulty settings—for example, recurring formations of play, timing asymmetries between players, or systematic differences in outcomes tied to camera perspective or interface feedback. Computational methods do not replace ethnographic description; they extend it, enabling the researcher to test intuitions, surface regularities, and situate local observations within broader distributions of play.

Moving beyond representation changes what counts as method as well as focus. The media systems that structure play—interfaces, controller mappings, camera logics, latency, matchmaking,

ranking, and feedback mechanisms—rarely announce themselves in interaction, yet they shape it continuously. In developing this orientation, we were influenced by work in the Chicago school, beginning with Robert E. Park's insistence that urban artifacts become institutional only insofar as they connect, through use and habit, to the "vital forces" of collective life (Park 1915). This attention to artifacts as lived infrastructure—tools that acquire social form through practice rather than design alone—carries forward in later work on infrastructure as relational, historically contingent, and often visible only at moments of breakdown or strain (Star and Ruhleder 1996; Bowker and Star 1999; Edwards 2003). To describe digital infrastructure, ethnography can draw on tools from platform and software studies, attending to how interfaces, affordances, and technical systems organize perception, action, and beyond what is visible at the interface (Larkin 2013; Gillespie 2018).

At this level, researchers might map interface elements and control schemes, documenting how different camera settings or controller configurations alter play, or comparing how the same actions register across platforms, game modes, or hardware setups. Researchers might deliberately vary settings, switch input devices, or repeat identical sequences of play under controlled conditions to observe how affordances shift. In some cases, this work extends to reading documentation, patch notes, or developer materials; in others, it requires forensic reconstruction through experimentation, reverse engineering, or systematic probing of system behavior. Social dynamics that appear interpersonal at the surface—cooperation, competition, reputation—often reflect the operation of ranking systems, feedback loops, or network effects that only become visible through such comparative and iterative analysis.

Regulatory constraints introduce a further methodological challenge. In the digital football game, enforcement occurs not only through explicit rules of play, but through automated systems: anti-cheat mechanisms, matchmaking thresholds, penalty regimes, content moderation, and account governance. These systems shape outcomes unevenly and often opaquely, operating at a distance from both player intention and situated interaction.

Our approach here was influenced by work that treats regulation not simply as formal law or stated policy, but as a distributed set of mediating practices that materialize through technical systems, routines of enforcement, and differential visibility (Clifford 1986; Rabinow 1977; Lessig 1999; Gillespie 2018). As with infrastructure, regulatory force often becomes ethnographically visible only in moments of intervention—through warnings, sanctions, exclusions, or sudden changes in access—rather than through everyday compliance. Attending to these moments allows ethnography to register how governance operates through code, platforms, and automated decision-making, even when it remains absent from the screen (Burawoy 1998).

Methodologically, this can involve tracking how identical actions receive different responses across accounts, matches, or time periods; documenting instances of warning, penalty, suspension, or exclusion; or comparing player experiences before and after rule changes or platform updates. Researchers might collect and analyze policy documents, community guidelines, patch notes, and enforcement notifications, while also attending to their practical effects as they register in play. Patterns of reward, restriction, or suppression often become visible only through longitudinal observation and comparison—by following accounts over time, contrasting sanctioned and unsanctioned behavior, or tracing how reputational and ranking systems modulate access and visibility. Regulation rarely appears solely on screen; its effects accumulate, shaping who can play, under what conditions, and which forms of action remain viable.

Case Study: Library Genesis

The layered approach outlined above emerges directly from our own ethnographic work. In what follows, we briefly revisit a concrete case study drawn from earlier research on Library Genesis,

one of the largest "shadow libraries" in the world, to show how digital ethnography in practice already requires movement across representation, media, and regulation.

Library Genesis (often abbreviated as Libgen) is a distributed digital library that aggregates and preserves millions of scholarly books and articles outside formal publishing and library infrastructures. Rather than treating the project as a site of piracy alone, our study approached it ethnographically—as a sociotechnical system composed of people, texts, software, and governance mechanisms, oriented less toward mass distribution than toward long-term preservation of knowledge (Yi Tenen and Foxman 2014). The analysis combined close qualitative observation with computational and infrastructural methods, making it a useful illustration of the approach sketched here.

At the level of representation, the study began with what was immediately visible: catalog records, metadata fields, file formats, interfaces for search and download, and traces of user interaction. We collected and analyzed bibliographic metadata at scale, examining patterns of duplication, classification, and absorption across collections. Computational analysis made it possible to identify how texts circulated, how collections grew, and how "canonical" packages emerged over time—patterns that could not be inferred from isolated observation alone.

Moving beyond representation required attention to media systems and software architecture. The organization of Libgen depended on specific technical choices: hashing schemes for de-duplication, database structures for indexing, forum software for coordination, and BitTorrent-based distribution for redundancy and resilience. Understanding how the library functioned meant reading documentation, examining code paths where possible, reconstructing workflows through forum archives, and tracing how infrastructural decisions shaped participation, authority, and access. What appeared as a single library interface in fact rested on a layered media ecology that structured what contributors could do and how the archive could survive.

Finally, the study confronted overlapping regimes of regulation. Library Genesis operated under community rules governing contribution, curation, and quality control, alongside corporate enforcement by publishers and state-level legal pressure that periodically reshaped the system's visibility and topology. These regulatory forces rarely appeared directly in everyday use, yet their effects accumulated over time—through takedowns, migrations, mirroring strategies, and shifts in governance. Tracing these dynamics required longitudinal observation, comparison across versions of the system, and attention to enforcement events as ethnographic data rather than external context.

Taken together, the Libgen case illustrates how digital ethnography can proceed in practice: by moving deliberately across surface representation, mediating infrastructures, and regulatory constraints, and by augmenting participant observation with computational analysis, media archaeology, platform study, and network analysis. Such surface representations can never exhaust the field, of course. Instead, the task of digital ethnography remains in assembling descriptions thick enough to register the conditions that make such appearances possible.

Dennis
Barbara
Linda
Olivia
Aya

Works Cited

Bowker, Geoffrey C., and Susan Leigh Star. 1999. *Sorting Things Out: Classification and Its Consequences*. Cambridge, MA: MIT Press.

Burawoy, Michael. 1998. "The Extended Case Method." *Sociological Theory* 16 (1): 4–33.

Clifford, James. 1986. "Introduction: Partial Truths." In *Writing Culture: The Poetics and Politics of Ethnography*, edited by James Clifford and George E. Marcus, 1–26. Berkeley: University of California Press.

Edwards, Paul N. 2003. "Infrastructure and Modernity: Force, Time, and Social Organization in the History of Sociotechnical Systems." In *Modernity and Technology*, edited by Thomas J. Misa, Philip Brey, and Andrew Feenberg, 185–225. Cambridge, MA: MIT Press.

Emerson, Robert M., Rachel I. Fretz, and Linda L. Shaw. 1995. *Writing Ethnographic Fieldnotes*. Chicago: University of Chicago Press.

Geertz, Clifford. 1973. *The Interpretation of Cultures*. New York: Basic Books.

Gillespie, Tarleton. 2018. *Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media*. New Haven, CT: Yale University Press.

Hargittai, Eszter. 2020. "Potential Biases in Big Data: Omitted Voices on Social Media." *Annals of the American Academy of Political and Social Science* 659 (1): 63–76.

Larkin, Brian. 2013. "The Politics and Poetics of Infrastructure." *Annual Review of Anthropology* 42: 327–43.

Lessig, Lawrence. 1999. *Code and Other Laws of Cyberspace*. New York: Basic Books.

Marcus, George E. 1995. "Ethnography In/Of the World System: The Emergence of Multi-Sited Ethnography." *Annual Review of Anthropology* 24: 95–117.

Malinowski, Bronisław. 1922. *Argonauts of the Western Pacific*. London: Routledge.

Park, Robert E. 1915. "The City: Suggestions for the Study of Human Nature in the Urban Environment." *American Journal of Sociology* 20 (5): 577–612.

Rabinow, Paul. 1977. *Reflections on Fieldwork in Morocco*. Berkeley: University of California Press.

Shaw, Aaron, and Benjamin Mako Hill. 2014. "Laboratories of Oligarchy? How the Iron Law Extends to Peer Production." *Journal of Communication* 64 (2): 215–238.

Star, Susan Leigh, and Karen Ruhleder. 1996. "Steps Toward an Ecology of Infrastructure: Design and Access for Large Information Spaces." *Information Systems Research* 7 (1): 111–134.

Yi Tenen, Dennis. 2017. *Plain Text: The Poetics of Computation*. Stanford, CA: Stanford University Press.

Yi Tenen, Dennis, and Maxwell Foxman. 2014. "Book Piracy as Peer Preservation." *Computational Culture* 4.