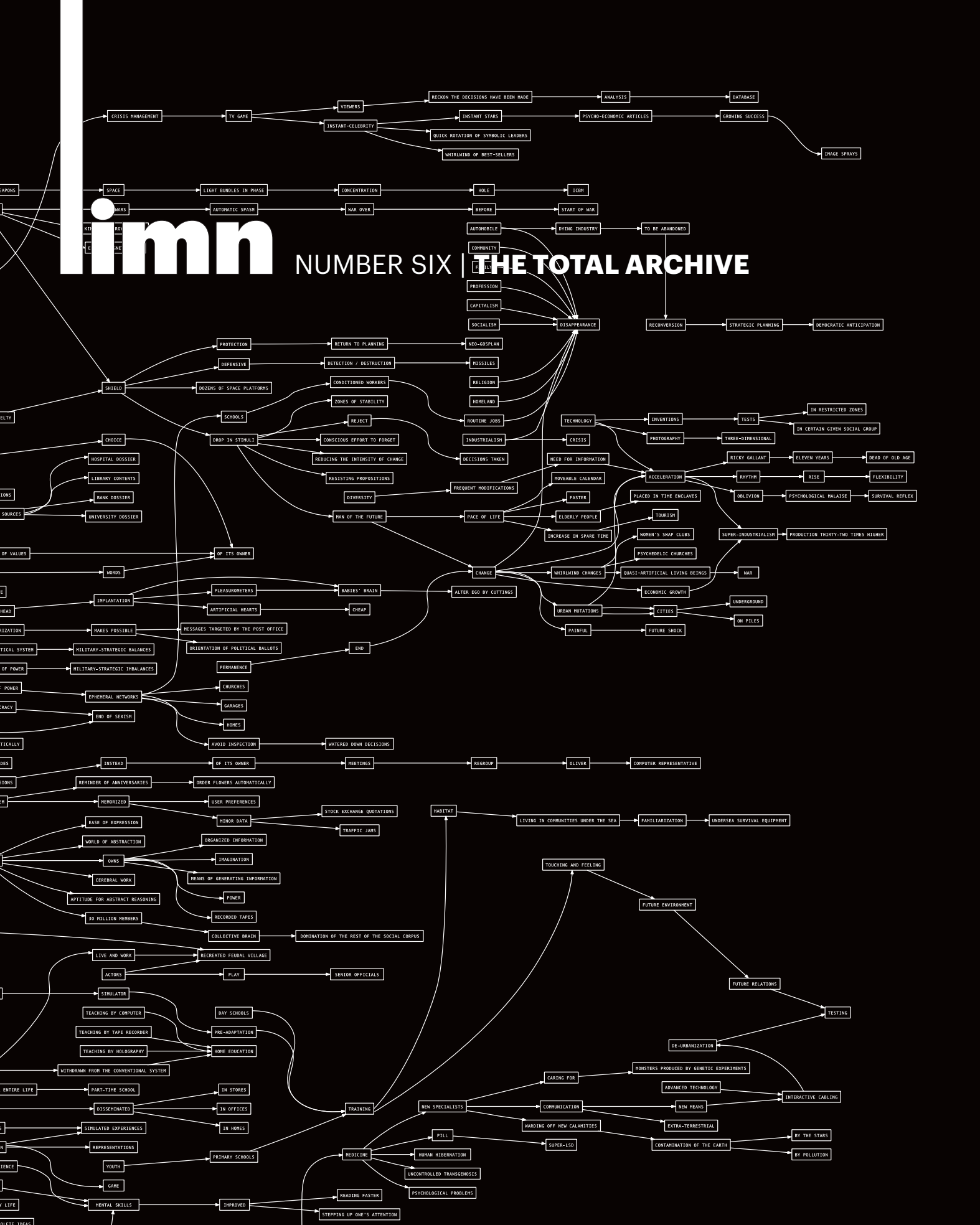

imn
NUMBER SIX | THE TOTAL ARCHIVE
CRISIS MANAGEMENT
TV GAME
VIEWERS
INSTANT-CELEBRITY
RECKON THE DECISIONS HAVE BEEN MADE
INSTANT STARS
QUICK ROTATION OF SYMBOLIC LEADERS
WHIRLWIND OF BEST-SELLERS
ANALYSIS
PSYCHO-ECONOMIC ARTICLES
DATABASE
GROWING SUCCESS
IMAGE SPRAYS
SPACE
LIGHT BUNDLES IN PHASE
CONCENTRATION
HOLE
ICBM
AUTOMATIC SPASM
WAR OVER
BEFORE
START OF WAR
AUTOMOBILE
DYING INDUSTRY
TO BE ABANDONED
COMMUNITY
PROFESSION
CAPITALISM
SOCIALISM
DISAPPEARANCE
RECONVERSION
STRATEGIC PLANNING
DEMOCRATIC ANTICIPATION
PROTECTION
RETURN TO PLANNING
NEO-GOSPLAN
DEFENSIVE
DETECTION / DESTRUCTION
MISSILES
SHIELD
DOZENS OF SPACE PLATFORMS
CONDITIONED WORKERS
RELIGION
HOMELAND
ZONES OF STABILITY
SCHOOLS
REJECT
ROUTINE JOBS
DROP IN STIMULI
CONSCIOUS EFFORT TO FORGET
REDUCING THE INTENSITY OF CHANGE
RESISTING PROPOSITIONS
DECISIONS TAKEN
INDUSTRIALISM
CRISIS
TECHNOLOGY
INVENTIONS
TESTS
IN RESTRICTED ZONES
IN CERTAIN GIVEN SOCIAL GROUP
PHOTOGRAPHY
THREE-DIMENSIONAL
CHOICE
HOSPITAL DOSSIER
LIBRARY CONTENTS
BANK DOSSIER
UNIVERSITY DOSSIER
DIVERSITY
FREQUENT MODIFICATIONS
NEED FOR INFORMATION
MOVEABLE CALENDAR
FASTER
MAN OF THE FUTURE
PACE OF LIFE
ELDERLY PEOPLE
INCREASE IN SPARE TIME
ACCELERATION
RICKY GALLANT
ELEVEN YEARS
DEAD OF OLD AGE
RHYTHM
RISE
FLEXIBILITY
OBLIVION
PSYCHOLOGICAL MALAISE
SURVIVAL REFLEX
PLACED IN TIME ENCLAVES
TOURISM
WOMEN'S SWAP CLUBS
PSYCHEDELIC CHURCHES
SUPER-INDUSTRIALISM
PRODUCTION THIRTY-TWO TIMES HIGHER
OF ITS OWNER
WORDS
CHANGE
WHIRLWIND CHANGES
QUASI-ARTIFICIAL LIVING BEINGS
WAR
ECONOMIC GROWTH
PLEASUROMETERS
BABIES' BRAIN
ALTER EGO BY CUTTINGS
IMPLANTATION
ARTIFICIAL HEARTS
CHEAP
URBAN MUTATIONS
CITIES
UNDERGROUND
ON PILES
MAKES POSSIBLE
MESSAGES TARGETED BY THE POST OFFICE
ORIENTATION OF POLITICAL BALLOTS
END
PAINFUL
FUTURE SHOCK
MILITARY-STRATEGIC BALANCES
MILITARY-STRATEGIC IMBALANCES
PERMANENCE
EPHEMERAL NETWORKS
CHURCHES
GARAGES
HOMES
END OF SEXISM
AVOID INSPECTION
WATERED DOWN DECISIONS
INSTEAD
OF ITS OWNER
MEETINGS
REGROUP
OLIVER
COMPUTER REPRESENTATIVE
REMINDER OF ANNIVERSARIES
ORDER FLOWERS AUTOMATICALLY
MEMORIZED
USER PREFERENCES
MINOR DATA
STOCK EXCHANGE QUOTATIONS
TRAFFIC JAMS
HABITAT
LIVING IN COMMUNITIES UNDER THE SEA
FAMILIARIZATION
UNDERSEA SURVIVAL EQUIPMENT
EASE OF EXPRESSION
WORLD OF ABSTRACTION
ORGANIZED INFORMATION
OWNS
IMAGINATION
CEREBRAL WORK
MEANS OF GENERATING INFORMATION
APTITUDE FOR ABSTRACT REASONING
POWER
30 MILLION MEMBERS
RECORDED TAPES
COLLECTIVE BRAIN
DOMINATION OF THE REST OF THE SOCIAL CORPUS
TOUCHING AND FEELING
FUTURE ENVIRONMENT
LIVE AND WORK
RECREATED FEUDAL VILLAGE
ACTORS
PLAY
SENIOR OFFICIALS
FUTURE RELATIONS
SIMULATOR
TESTING
TEACHING BY COMPUTER
DAY SCHOOLS
TEACHING BY TAPE RECORDER
PRE-ADAPTATION
TEACHING BY HOLOGRAPHY
HOME EDUCATION
DE-URBANIZATION
WITHDRAWN FROM THE CONVENTIONAL SYSTEM
MONSTERS PRODUCED BY GENETIC EXPERIMENTS
CARING FOR
ADVANCED TECHNOLOGY
INTERACTIVE CABLING
ENTIRE LIFE
PART-TIME SCHOOL
IN STORES
DISSEMINATED
IN OFFICES
IN HOMES
TRAINING
NEW SPECIALISTS
COMMUNICATION
NEW MEANS
SIMULATED EXPERIENCES
WARDING OFF NEW CALAMITIES
EXTRA-TERRESTRIAL
REPRESENTATIONS
PILL
SUPER-LSD
CONTAMINATION OF THE EARTH
BY THE STARS
BY POLLUTION
YOUTH
PRIMARY SCHOOLS
MEDICINE
HUMAN HIBERNATION
GAME
UNCONTROLLED TRANSGENOSIS
PSYCHOLOGICAL PROBLEMS
MENTAL SKILLS
IMPROVED
READING FASTER
STEPPING UP ONE'S ATTENTION

LIMN NUMBER SIX **THE TOTAL ARCHIVE**

Edited by Boris Jardine and Christopher Kelty

Preface

Archives make the future. Editors **Boris Jardine** and **Christopher Kelty** explore how archives govern us.

VAST ACCUMULATIONS of data, documents, records, and samples saturate our world: bulk collection of phone calls by the NSA and GCHQ; Google, Amazon or Facebook's ambitions to collect and store all data or know every preference of every individual; India's monumental efforts to give everyone a number, and maybe an iris scan; hundreds of thousands of whole genome sequences; seed banks of all existing plants, and of course, the ancient and on-going ambitions to create universal libraries of books, or their surrogates.

Just what is the purpose of these optimistically total archives – beyond their own internal logic of completeness? Etymologically speaking, archives are related to government—the site of public records, the town hall, the records of the rulers *(archons)*. Governing a collective—whether people in a territory, consumers of services or goods, or victims of an injustice—requires keeping and consulting records of all kinds; but this practice itself can also generate new forms of governing, and new kinds of collectives, by its very execution. Thinking about our contemporary obsession with vast accumulations through the figure of the archive poses questions concerning the relationships between three things:

1. the systematic accumulation of documents, records, samples or data;
2. a form of government and governing; and
3. a particular conception of a collectivity or collective kind.

What kinds of collectivities are formed by contemporary accumulations? What kind of government or management do they make possible? And who are the governors, particularly in contexts where those doing the accumulation are not agents of a traditional government?

This issue of *Limn* asks authors to consider the way the archive—as a figure for a particular mode of government—might shed light on the contemporary collections, indexes, databases, analytics, and surveillance, and the collectives implied or brought into being by them. At the very least, we have demanded more precision than is found in breathless mainstream media accounts of big data: Is a database an archive? When is an index a database? How is a collection of paper records different when it comes with tiny spot of dried blood on it than when it does not? What differences make a difference when we talk about a seed bank instead of a seed database or a repository of open access documents instead of a pirate library of scanned books? Indeed, are digital collections 'archives' at all—do they undermine the existence of archives traditionally conceived?

There are limitations to thinking of contemporary vast accumulations as archives: they often lack a single point of authority or intentionality. Rather than a government office, a corporate archivist, an individual collector, they span all these things—data is shared, bought and sold; samples are shipped, frozen, reproduced and mutated; digital records are collected and modified automatically or indiscriminately using procedures and algorithms that sort and filter, often in ways that trigger immediate and consequential action—from terrorist watchlists to mandatory newborn screening.

It's not the size or the comprehensiveness of contemporary accumulations that makes them different. Archives—like the **Cairo Genizah**—have always been big and messy. The Cairo Genizah began with just a *rule*—an injunction to preserve any holy document, regardless of purpose. This resulted in a form of "automatic" collection—an ancient logfile, as it were, never deleted, rarely consulted, but containing valuable and no doubt "private" information of all kinds. Repurposed as a source, this archive now determines what we know about a whole era of human history, throughout much of the Mediterranean and Middle East. But did the people who used it agonize about what went into it? Did they structure their understanding or their behavior according to who might consult it: the state, historians of the future, G-d perhaps?

Similarly, archives have never been stable, unchanging supplements to government, or perfect reflections of collectives: they are dynamically constitutive of those collectives and their government. Information enters archives, but it also escapes them. The perfect total archive would leave no question unanswered, no gene unsequenced, no seed unsaved, no phone call unheard, no book unread, uncatalogued or uncited. But such perfections exist only in the fictions of Jorge Luis Borges, who, Kate Hayles **reminded us**, provides the limit imaginations of the archive—the Aleph and the Library of Babel.

Instead, every archive is partial, and every partial archive has its anxieties: incompleteness, redactions, mis-filings, duplications, obfuscations, ignorance, secrets. The dream of total archives governing perfectly in Borges is interrupted by the reality of total governance with imperfect archives. Like a government built around the concept of territory—with all its porous boundaries, shifting fences and walls, and undefended hinterlands—a government built around the database or archive encounters leaks, breakdowns, shifting technologies and ineffective firewalls. The case of "**Digital India**" shows that the simple problem of *duplication* (and the techniques of de-duplicating) troubles the system. Every attempt to combat corruption reveals new possibilities for it. Multiple databases raise the question of how to live with multiple sovereigns, or navigate between them.

Where then, does the desire for totality come from? What forms of government, and what kinds of collectives demand totality—even if a Phyrric form—and what kinds resist it? The Cold War emerges in the essays collected here as a particularly fertile ground for the accumulation of masses of data, and as a key site for understanding the contemporary legacy of that archival urge. Total archives from this era span every discipline. It is the time when Alan Lomax **(Laemmli)** developed the study of "Choreometrics" and travelled the world in pursuit of an archive of all bodily movement; it is the time when anthropologists and psychologists could imagine a "database of dreams" that would record the inner lives of people around the world—but which ultimately became

partial pastiche instead of total archive (**Lemov**, p. 30). From this era comes the story of the "bombing encyclopaedia" that sought to catalogue every bombable target the US military might conceivably attack, and which now forms the basis for a generic technique of "catastrophe modelling" used in finance, disaster planning and disease surveillance today (**Collier and Lakoff**, p. 53). From the same period comes the story of a humble social science device: Cantril's Ladder, used to measure happiness globally and longitudinally ("where on this ladder would you place yourself?"). Today it forms the basis of a globally significant economic indicator and measure of well-being—Gross National Happiness (**Jardine**, p. 48). The mid-century also saw an increase in the scientific collection of language data and blood group typing for transfusion, both of which were archives not just for preservation, but developed in order to promote, defend and study diversity around the world (**Bangham**, p. 43; **Kaplan**, p. 64). Both form the basis for new forms of biological and linguistic knowledge production (**Reardon**, p. 72).

All these cases from the past demonstrate that archives are never just about representation or preservation—they also perform, create, and remake collectives. They participate in governing just as much as they represent some reality or object of study. But these mid-century modern archives seem somehow quaint and controllable by comparison to those of today. Today it sometimes seems that collecting *everything* is just what we do. Why not log it? Why not keep a copy? Why not digitize it? Why not store—and share, analyse or leak—a petabyte of data every 48 hours? Costs fall and digital information properly structured can produce its own traces automatically. The ability to accumulate has outstripped the cost of or need for doing so by leaps and bounds—but unlike the Genizah, such accumulations are not singular or hidden, but duplicated, ramified, leaked and regularly consulted.

Today we can automatically log every transaction ever conducted in a particular currency—and Bitcoin is just such a global experiment. It has been imagined by a technologically sophisticated network of true believers who see not just a new currency, but a total system of governance by ledger, a transformed concept of "contracting" in which trust (interpersonal and in a sovereign) is replaced by math (**Brunton**, p. 83 and p. 87). Today we can imagine a perverse census at a shocking scale: two competing efforts to give everyone in India a unique number—or maybe two unique numbers—stored in supposedly safe databases whose purposes include everything from combatting corruption to delivering "services" to remaking the very composition of the collective (**Cohen**, p. 77). We can earnestly aim at collecting "All the World's Knowledge" in an online encyclopaedia *including* records of everything we've deleted or shouldn't have kept in the first place (**Binns**, p. 11). And it is not just us, but the information itself that is governed today: the very demand for "open access" to all the world's scholarship turns out to be more legitimately the work of activist-scientists schooled in the tradition of samizdat publishing than it does of a corporate giant like Google, for whom a vast accumulation has turned out to be an incredible liability (**Bodó**, p. 19; **Murrell**, p. 15).

The drive to collect everything simultaneously produces anxieties of surveillance and elaborations of vitality: at one and the same time we fear the forms of government of new and old collectivities being rendered possible by our accumulations and we insist on the impossibility of its power. We decry surveillance and intrusion, but we say: that body made of data is not me—I cannot be represented by a database no matter how total. And yet, I simply cannot function without it. The idea that the practice of governing might change in response to the availability of information is different from the accusation that power desires total information. The conspiracy theories by which privacy and surveillance activists attribute to government an unchanging desire to hoard and make secret implies a kind of sovereign power that is only part of how we govern through accumulation. But accumulations can also govern by producing new forms of discipline (**Bowker**, p. 40; **Poleykett et. al.**, p. 26), as well as by providing new resources for contestation, satire, resistance, or sabotage (**Balasz**, p. 19; **Previeux**, p. 4). They mutate the forms of government available to everyone, but never just in the way intended, and certainly not equally.

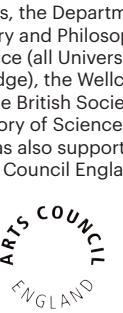
Alongside the questions of governance and collectivities, this issue of *Limn* also stages the question of the aesthetics of the archive. Much of this work dramatizes the elaboration of vitality that accumulations can produce: the **UA artists** group explores different ways to index temporality, disconnection, desires both for the total, and ways to escape it. The mysterious **Valaco Archive** explores and extends the limit conditions that we know from Borges stories through the archives of a single, real (?) person. **Fabienne Hess** most directly subjects herself to the total archive, and to the seemingly infinite expanse of tiny variations. And **Julien Previeux** dramatizes the lament of loss and partiality by collecting books that have been cut loose from the total archive, and graphically recovering the knowledge they contain. In an only half-joking way, he asks the question of the issue: what is this knowledge made from? Does it govern us now? Did it in the past? Will it in the future? ■

Boris Jardine and Christopher Kelty
March 2016

limn is published as needed. Issue #6 is edited by **Boris Jardine** and **Christopher Kelty**. This issue is set using Christian Schwartz' Graphik, Dino dos Santos' Leitura and Velino Poster, and Hoefler & Co.'s Whitney typefaces. Layout by **Martin Hoyem/American Ethnography.** The General Editors of Limn are **Stephen J. Collier, Christopher M. Kelty,** and **Andrew Lakoff.**

Publication assistance also provided by the Research Cluster in Science, Technology and Society at the University of Southern California. More at http://limn.it/

|| Some of the essays in this issue of Limn draw on material first presented at the Total Archive Conference, University of Cambridge, 19/20 March 2015. The conference was convened by Boris Jardine, Matthew Drage and Ruth Horry, and was supported by grants from the Centre for Research in the Arts, Social Sciences and Humanities, the School of Humanities and Social Sciences, the Department of History and Philosophy of Science (all University of Cambridge), the Wellcome Trust, the British Society for the History of Science. This issue was also supported by the Arts Council England.

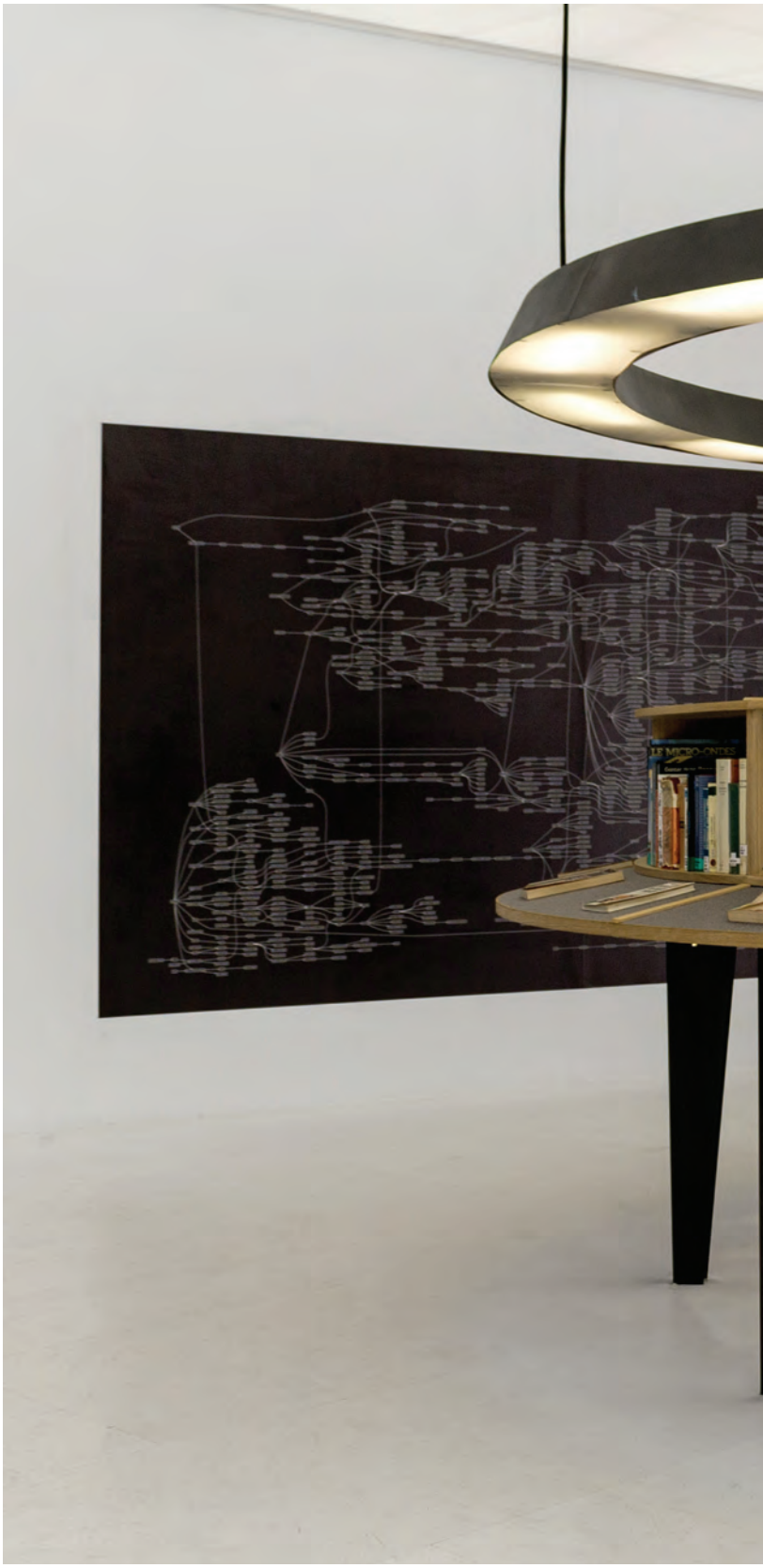


The totality of true propositions (before) (2008-2009)

LIKE A CARTOGRAPHIC EXERCISE, **JULIEN PRÉVIEUX** TRACES THE OUTLINES OF A COMPLETELY UCHRONIC PARALLEL FUTURE, NOT WITHOUT WIT.

Julien Prévieux's artwork *The Totality of True Proposition (Before)* (2008–2009) is a bookcase surrounded by huge diagrams. The bookcase contains numerous books and user's guides that have been deemed obsolete by French librarians and slated for destruction. Public library space is often limited, and new books replace the old ones. Prévieux's work is the result of a lengthy task of collecting outmoded books from public library collections, bringing together manuals and handbooks such as old dictionaries (like *Le Nouveau Petit Larousse*, published in 1959) and computer books (like *Windows 95 for Dummies*), as well as historical and theoretical volumes such as *USSR, The Country Where The Sun Never Goes Down* (1971) by Emil Schulthess and *The Modern Warfare* (1984) by William V. Kennedy, whose modernity and ideas have not survived the inexorable passage of time. Overlooked, scorned, and on the sidelines of state-of-the-art knowledge, these books, once reorganized in a library of linguistic, technical, and historical puzzles, contain knowledge that is no longer germane, but still makes sense on an ironic and poetic level. A particular section of this *knowledge freezer* contains books that forecast what tomorrow would be, such as *Future Shock*, the famous bestseller written by Alvin Toffler in 1970. Written 20 or 30 years ago, the content of these predictive books sound offbeat if we compare them with what we experience today. Julien Prévieux traced huge diagrams by means of data mining, transforming the themes of these works into a set of crazy oracles. Like a cartographic exercise, it traces the outlines of a completely uchronic parallel future, not without wit.

See more of Julien Prévieux's work at:
http://www.previeux.net/



Ben Outhwaite tells the stories of the people who immerse themselves in one of the most valuable total archives in existence—the Cairo Genizah.

A HOARD OF HEBREW MSS

THE CAIRO GENIZAH IS AN ACCUMULATION OF 1,000 YEARS OF Jewish culture in the lands of Islam. From the time of its discovery in the Ben Ezra Synagogue in Fustat, Old Cairo (or rediscovery, since the storeroom from which it was finally removed in the late nineteenth century had been peered into by intrepid scholars, treasure-hunters, and tourists over several centuries), the collection's sheer size has been guessed, estimated, and then, finally, counted in the early years of the twenty-first century.[1] The Taylor-Schechter Cairo Genizah Collection in Cambridge University Library, the largest single grouping of Genizah Collection documents by an order of magnitude, boasts an inventory in the form of an Excel spreadsheet (enumerating only the class marks of the manuscripts) that itself is a whopping 32 megabytes big.[2] A total of 137,000 class marks delineate 193,000 Genizah fragments: in some cases these fragments are tiny scraps, with only a few Hebrew or Arabic letters still visible, but "fragment" also describes broadsheet-sized marriage deeds of once luxuriously white parchment or carpet-sized sections of liturgically correct Torah scroll, tanned to a deep, rich brown. In size and significance, the Collection dwarfs its nearest rival, the Dead Sea Scrolls, and in its extent and diversity only that other great Egyptian collection, the

1 Solomon Schechter, in his first description of the discovery in the Times (1897), states "The number of fragments procured by me equates, I think, to about 40,000." By any manner of counting he was way off.

2 The inventory, carried out by the staff of the Genizah Research Unit in Cambridge, was the brainchild of Professor Yaacov Choueka of the Friedberg Genizah Project, who recognized that to digitize the entire collection, we should first and finally know its full extent.

PREVIOUS PAGE FROM TOP: Unsorted arabic fragments; Schechter at Work in Cambridge.

ABOVE: Catalogued fragments from the Cairo Genizah.

ABOVE: Shelomo Dov Goitein (1900-1985).

Oxyrhynchus papyri, comes close.

The emergence of the Cairo Genizah from its dark Fustat storeroom has often been told as an exhilarating tale of late-Victorian derring-do, orientalist travelogue, and amazing serendipity.[3] Solomon Schechter, the Cambridge scholar who brought the collection out of Egypt in 1896–1897 through perseverance, charm and—thanks to his Cambridge friend Charles Taylor's money—bribery, described it most elegantly:

> The Genizah, to explore which was the object of my late travels in the east, is an old Jewish institution. The word is derived from the old Hebrew verb "ganaz," and signifies treasure-house or hiding-place. When applied to books it means much the same thing as burial means in the case of men. When the spirit is gone, we put the corpse out of sight to protect it from abuse. In like manner, when the writing is worn out, we hide the book to preserve it from profanation. The contents of the book go up to heaven like the soul (Schechter 1897).

Schechter's name, alongside that of the money-man, Taylor, is forever attached to the collection in Cambridge. Generously donating it to the University Library, he found a higher calling and a more prestigious role at New York's Jewish Theological Seminary. Before his death in 1915 he had made a number of impressive discoveries in the Genizah, but he recognized himself that he had barely scratched the surface. Schechter fished for texts that interested him: the monumental task of retrieving the collection done; he could afford an academic dilettantism thereafter. He understood that it was a collection of a scale and scope beyond one scholar, a combination, as he put it, "of sacred lumber-room and secular record office" that would prove "interesting alike to the theologian and the historian" (Schechter 1897). But at a stroke, the world of Jewish studies was changed forever.

The Genizah has an archaeological character. Material seems to have been deposited as a result of different processes, or by the application of different criteria, throughout the history of the Jewish community of Fustat, piling up in stratified layers. Dig deep, and at the root of it all is the rabbinic prohibition against defiling the name of God: a religious text cannot be aimlessly discarded, but must be carefully stowed away. And so, at the base of the collection, we find there tens of thousands of leaves from the Hebrew Bible, with its countless repetitions of the Tetragrammaton: the sacred, ineffable *Yod-He-Waw-He* that spells the name of God. To these are added vast numbers of pages from prayer books, which attest to medieval Judaism's wide and varied designations for the deity and his different facets: "the Place," "the Rock," "He who Spoke." The earliest stratum of the Genizah, material written before the tenth century (as far back as the fifth or sixth century if we count the underwriting of palimpsests), is just as we might expect: Bibles, prayers, and other theological works of a traditional type.

Today, however, the Genizah Collection is as famous for its documentary material—Schechter's "secular record office"—as for its Bibles and liturgies. The earliest documents probably owe their survival not to deliberate consignment into the Genizah, but to being caught up among other papers—perhaps those of a deceased Jewish community leader—and

3 The tale is beautifully told in Hoffman and Cole (2011) but there is much of interest in S. C. Reif (2000) and M. Glickman (2010).

ABOVE: Catalogued fragments from the Cairo Genizah. **RIGHT:** Drawing of Ben Ezra Synagogue.

ABOVE: Babylonian Gaon Nehemiah ha-Kohen's ca. 960 CE epistle from the Academy of Pumbeditha in Iraq

deposited, unsorted, into the Genizah chamber. The Babylonian Gaon Nehemiah ha-Kohen's ca. 960 CE epistle from the Academy of Pumbeditha in Iraq, a big fancy letter on expensive paper with—uniquely—a geonic bulla still attached, is probably one such example.[4] But by the beginning of the eleventh century, the deposit of documentary archives appears to have become commonplace, as letters, legal documents and accounting records all pile up in the thousands. Whether it is from the common use of divine names in the correspondence or the perceived sacred character of the Hebrew script itself (*lešon ha-qodeš*, or "the Holy Language"), is unclear, but nonetheless the Fustat community clearly felt a great need to preserve their written legacy.

The historian Alexander Marx stated matter-of-factly, and a little underwhelmingly, in his postwar summation quite how important this documentary archive was: "The period from the tenth century to the twelfth century, which was among those shrouded in darkness in Jewish history, has become known to us in considerable detail" (Marx 1946/47:186). The Genizah research of early scholars such as Cowley, Lewin, Assaf, Chapira, and Jacob Mann painted our first pictures of the political history of the Jews under Islam. Mann dipped into the boxes of manuscripts in Cambridge University Library and revived the little-known Jerusalem Academy, gave us its rites, its tradition of reading the Pentateuch in a three-year cycle, and examined the tussles between the Palestinian and Babylonian Jewish communities in Fustat. But his work is by its nature patchy, as he dug down into a vast, uncataloged, and little-understood collection. Though he made attempts at a narrative history, the reader can clearly discern that Mann is forced to segue from one episode to another in a succession of vignettes, snatches of Jewish history frozen in time. As Hebrew, which Mann knew well, gave way to Judaeo-Arabic in the letters and deeds that he relied on, Mann's picture of Jewish history becomes correspondingly more attenuated.[5]

The Fustat hoard found its chronicler

4 Cambridge University Library Jacques Mosseri Genizah Collection, Mosseri VIII.479. The Mosseri Collection is another Cairo Genizah Collection in Cambridge with an interesting and, ultimately, ironic history. See Outhwaite (2009).

5 Hebrew was still the preferred language of most official Jewish correspondence in the early eleventh century, but it was superseded by Judaeo-Arabic following the decline of Jewish life in Palestine in the second half of the century. Mann didn't read Arabic, and his two volumes of *The Jews in Egypt and in Palestine under the Fāṭimid caliphs; a contribution to their political and communal history based chiefly on genizah material hitherto unpublished* (2 vols; London etc, 1920–22) contain almost no Judaeo-Arabic sources. For the history of documentary Hebrew in this period, see Outhwaite (2013).

in the 1950s with S. D. Goitein. Goitein's interest in the social and economic history of the eleventh to thirteenth centuries put flesh on the skeleton assembled by Mann and others, and culminated in a masterpiece, his five volumes (plus a sixth index volume) of *A Mediterranean Society*, a work that has defined the size and shape of the documentary archive for a whole generation of subsequent scholars. Goitein, an Arabist, linguist, and anthropologist by training, transformed himself—or was he shaped by the Genizah?—over a whole series of monographs and hundreds of articles as he threw himself into the world of the Genizah. Delighting initially in the economic texts, which threw light, for instance, on the price of bread and the routes of the wheat trade over the course of hundreds of years, Goitein left no detail of Mediterranean commerce unmentioned in Volume One of *A Mediterranean Society*. Subtitled "Economic Foundations," it has a whole chapter on packing and packaging, down to descriptions of the type of wickerwork baskets used to carry sal ammoniac.

In Volume Two Goitein moved on from the economics of the Mediterranean world and began to examine, in his usual detail, the politics and administrative practices of the Jewish institutions of Fustat and their relations with the Islamic court. In chapters such as "Communal Organization and Institutions" and "Interfaith Relations," Goitein revealed the workings of Egyptian politics and brought to life mad viziers, corrupt beadles, and other splendid Genizah archetypes. But with all this time spent immersed in the daily lives of Fustat's citizens, fascinated by the casual detail they inserted in their letters and other documents and the ease with which they mixed business and personal affairs, Goitein's natural inclination to anthropological research was awoken, and by Volumes Three and Four ("The Family" and "Daily Life") he was a full-blown social historian, with intimate digressions on "At the dressing table" and "Fumigating and freshening" as he nosed his way into his subjects' courtyards, their houses and even their bedrooms. By Volume Five Goitein's evolution was complete: "The Individual: Portrait of a Mediterranean Personality of the High Middle Ages as Reflected in the Cairo Geniza." Goitein had become the Genizah specialist par excellence, the socioeconomic historian with a messianic fervor for the archive, capable of peering even into the heads of Fustat's Jewish citizenry. Far removed from his laborious listings of commodities and professions in volume one, in Volume Five he rises to poetic heights:

> It is this touch of an expanded existence which makes reading the majority of the Genizah correspondence so pleasant. It was a civilized world, of people who knew how to behave, who were considerate, paying proper attention to their fellowmen. He knew his station in society but was not unduly deferential to his superiors. To be sure, the very cadence of Arabic speech sounds hyperbolic to our ears. But, in general, as befitting busy people, their statements were concise, clear, and to the point – somehow reflecting the clarity of the Mediterranean sky (Goitein 1967–1993:7).

Unlike Solomon Schechter, who was able to walk away—indeed give away the Genizah—and start an altogether different career in America, Goitein immersed himself, like a romantic orientalist, into the world which he studied. Mann, never quite coming to grips with the sources, remained apart, detached, as if looking in from a distance, an astronomer with an imperfect telescope, seeking to make sense of only partial information. Goitein, however, metaphorically donned native garb and haunted the souks and alleyways of medieval Fustat, leaving nothing unread or unreported and relying on his imagination to fill the gaps. No one has better interpreted the everyday texts of that lost world, or squeezed so much life out of discarded ephemera. But no other collection has had such a seductive effect. The unique combination of its immense size but remarkable, detailed intimacy gives the Cairo Genizah an exhilarating, neverending attraction, to the point of scholarly addiction. ■

DR. BEN OUTHWAITE *has been head of the Genizah Research Unit in Cambridge University Library since 2006.*

BIBLIOGRAPHY

Glickman, M. 2010. *Sacred Treasure—The Cairo Genizah: The Amazing Discoveries of Forgotten Jewish History in an Egyptian Synagogue Attic*. Woodstock, VT: Jewish Lights.

Goitein, S. D. 1967–1993. *A Mediterranean Society: The Jewish Communities of the Arab World as Portrayed in the Documents of the Cairo Genizah*. 5 vols. with an index volume by P. Sanders. Berkeley, CA: University of California Press.

Hoffman, A., and Cole, P. 2011. *Sacred Trash: The Lost and Found World of the Cairo Geniza*. New York: Schocken.

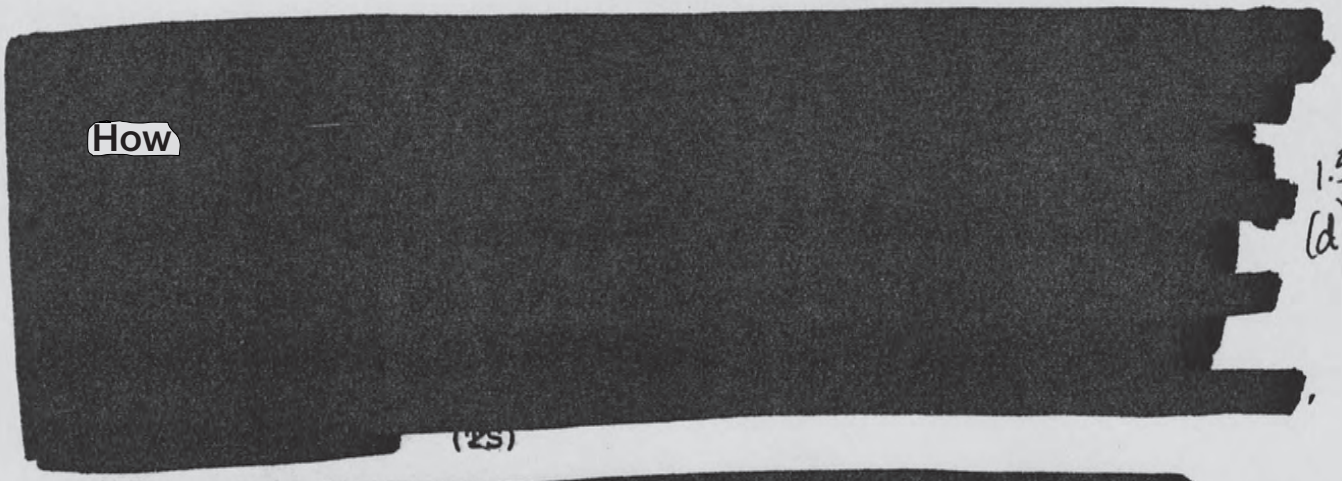
Marx, Alexander. 1946/47. "The Importance of the Geniza for Jewish History." *Proceedings of the American Academy for Jewish Research* 16:183–204.

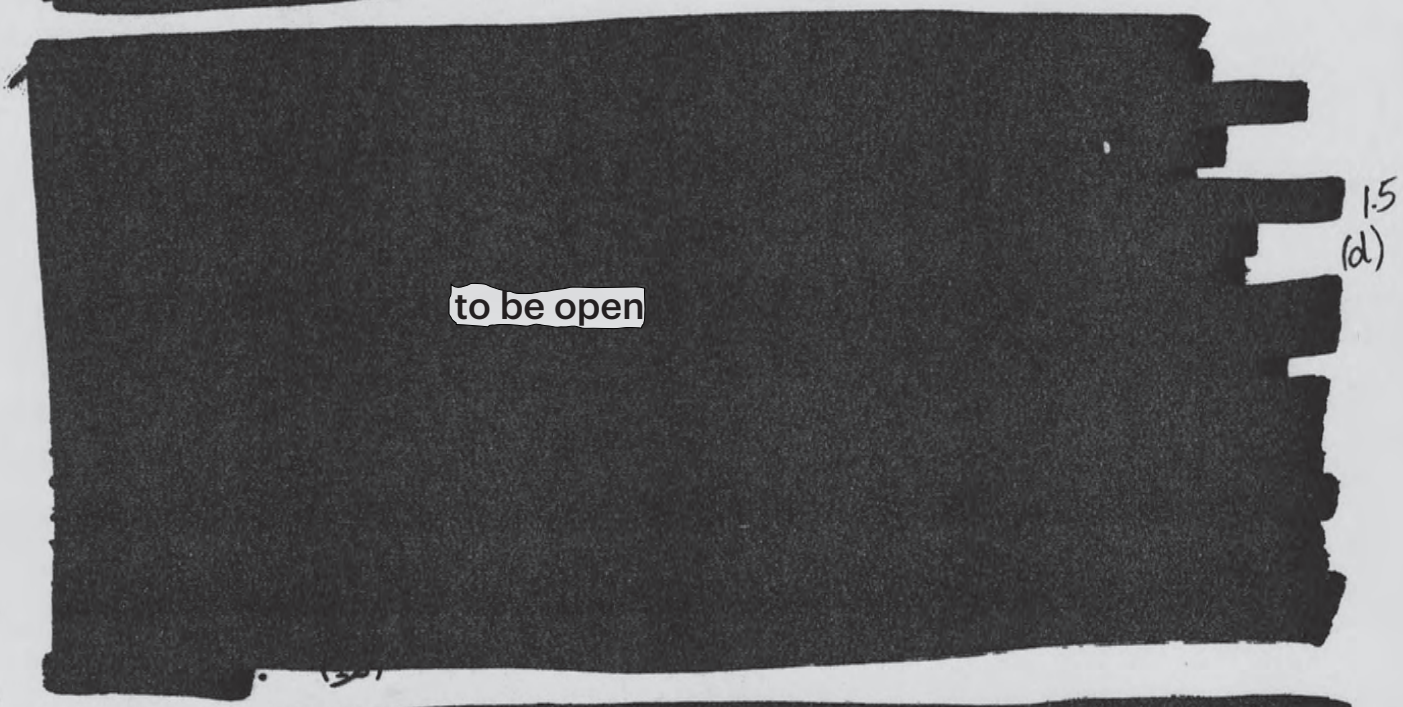
Outhwaite, B. 2009. "Taylor, Schechter and Jacques Mosseri." In *Charles Taylor and the Genizah Collection*, edited by Stefan Reif (pp. 75–80). Cambridge, UK: St John's College.

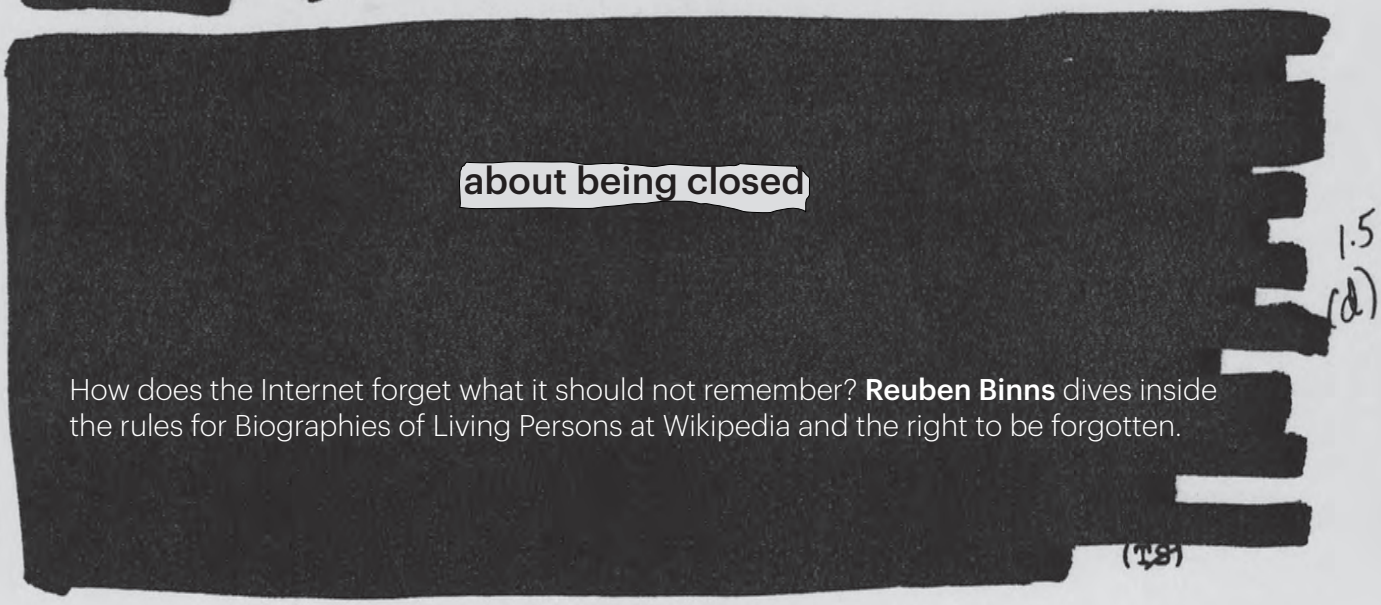
———. 2013. "Lines of Communication: Medieval Hebrew Letters of the 11th Century." In *Scribes as Agents of Language Change*, edited by E-M. Wagner, B. Outhwaite and B. Beinhoff (pp. 183–198). Berlin, Germany: Walter de Gruyter Inc.

Reif, S. C. 2000. *A Jewish Archive from Old Cairo: The History of Cambridge University's Genizah Collection*. Richmond, UK: Routledge.

Schechter, Solomon. 1897. "A Hoard of Hebrew MSS." *The Times*, August 3.







How does the Internet forget what it should not remember? **Reuben Binns** dives inside the rules for Biographies of Living Persons at Wikipedia and the right to be forgotten.

IN THE SUMMER OF 2014, WIKIPEDIA CREated a public list of pages that have been partly hidden from search engines for privacy reasons (Wikimedia Foundation 2015). It includes entries on criminals, famous musicians, and a chess player, all of whom appear to have made requests to Google to have the content de-listed from searches on their name. This "index of the de-indexed" is one of the many curious by-products of the online encyclopedia's ongoing construction. It encapsulates a peculiar set of contradictions arising out of the project's concurrent imperatives: to be at once selective and comprehensive; and to both exclude the vast majority of edits while also maintaining its radical openness.

The list, "Notices received from search engines," comprises links that have been removed from certain search engine results under European data privacy laws. The so-called "Right to Be Forgotten," established by a European Court of Justice ruling in May 2014, confirmed that European citizens have the right to request the removal of links to certain content about them when their name is entered into a search engine.[1] The right has its legal basis in decades-old data protection laws, but had been unenforced until a Spanish court ruled in favor of an individual who wanted Google to remove a link to a news article detailing his previously unpaid debts. The court agreed that Google would be required to remove the link to the article from search results that were based on the individual's name. This opened up the floodgates for other individuals to make similar requests under what became known as the "right to be forgotten."

As many commentators have noted, the phrase "the right to be forgotten" is misleading; the original content isn't removed, but only becomes harder to find using the individual's name. In each case the search engine is required to weigh the values of privacy and the public interest before accepting or rejecting a request. If Google's in-house arbitrators do approve a request to de-list, they notify the website that has been de-listed.

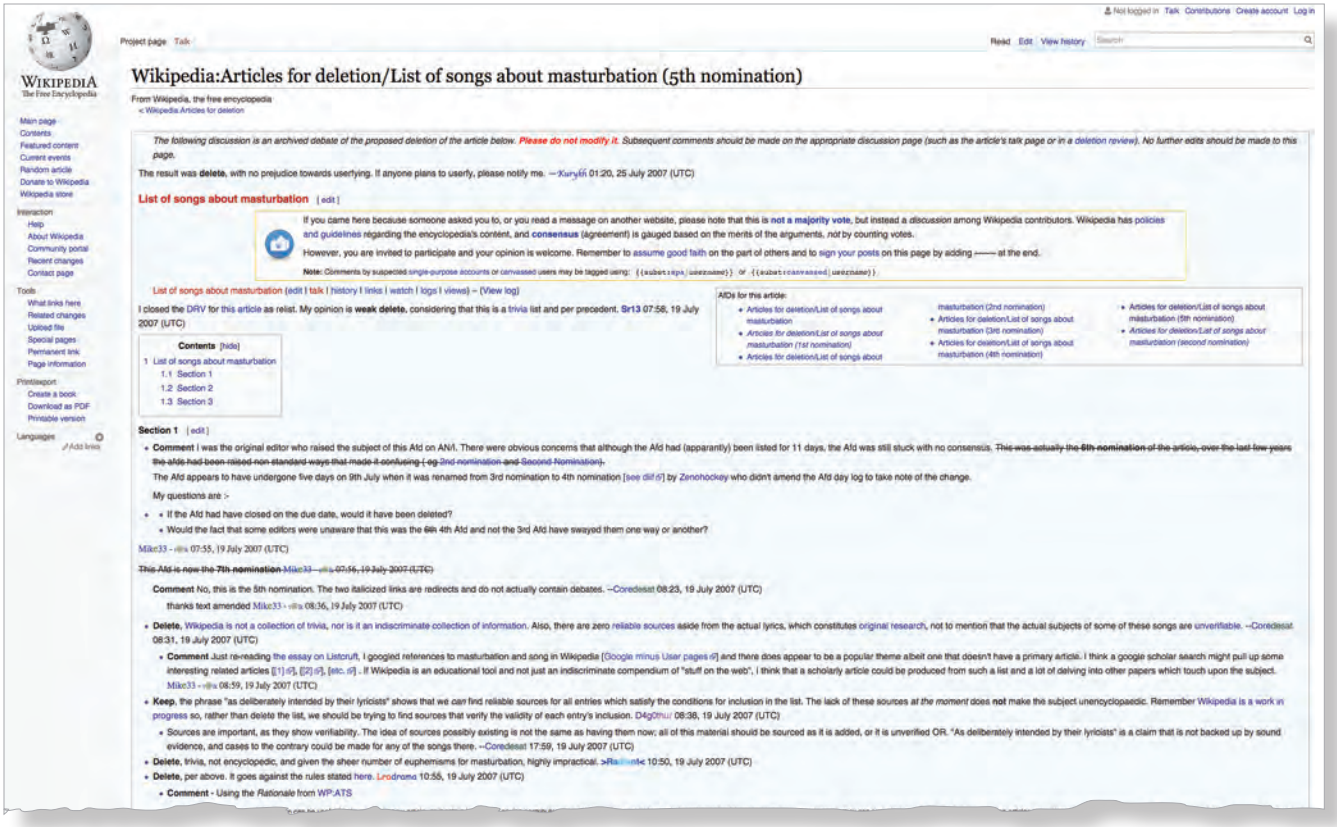
The Wikimedia Foundation, the nonprofit that operates Wikipedia, compiles such notices in the index. Clearly, publishing a list of pages containing the personal information of subjects who have explicitly attempted to obscure them somewhat undermines the purpose of the right to be forgotten. In this sense, the index of the de-indexed looks like a retaliatory blow struck by defenders of openness in their battle against censorship and undeserved privacy. Wikimedia's press release accompanying the index supports this interpretation; it argues that content "should not be hidden from Internet users seeking truthful and relevant information," and that the ruling "runs counter to the ethos and values of the Wikimedia movement" (Wikimedia 2014). Wikimedia is not alone in making these notifications public; the BBC also maintains a list of affected pages to preserve "the integrity of the BBC's online archive" (McIntosh 2015).

We could see this simply as a clash between those who think certain information about individuals should be made public and those who don't. But Wikipedia's existing processes for handling the deletion and selection of content suggest a more nuanced position. The apparent hostility toward removing information on grounds of privacy belies the measured stance to be found in Wikipedia's long-established policies. For many years prior to Google Spain, the project has had its own policy on biographies of living persons (BOLP). This includes many admirable principles that echo those laid out in the court decision. Biographies should be based on up-to-date and reputable sources, "relevant to a disinterested article about the subject," with due regard for privacy. It warns against spreading "titillating claims about people's lives; the possibility of harm to living subjects must always be considered when exercising editorial judgment" (Wikipedia, 2015). Furthermore, biographies may in some cases be entirely *deleted*, upon request, if the individual concerned is a relatively unknown, non-public figure.

These rules have themselves been developed in the "wiki-way"—through online discussion and consensus building—and aim to balance various criteria, including the public interest, privacy, and freedom of speech. Compare the requirements of the BOLP to the considerations outlined in the Google Spain decision, and they begin to look roughly equivalent (in some cases, the BOLP appears to impose an even stronger imperative to forget). So in addition to the recent list of pages de-indexed by search engines, there is a much older record of changes made by the site's editors in accordance with its own self-imposed privacy principles outlined in the BOLP. Some recent examples of privacy-motivated deletions that have arisen out of this policy include removing a link between an author's real name and a suspected pen name; removing contextual information about an individuals' family members; and removing references from a medical doctor's biography to rumors that his or her medical license had been revoked.

Given the substantially overlapping criteria between the right to be forgotten and the BOLP policy, why would the Wikimedia Foundation denounce the former while implicitly endorsing the latter? The notion that this controversy is simply

1 Google Spain SL, Google Inc. v Agencia Española de Protección de Datos, Mario Costeja González (2014)



AN EXAMPLE: From the discussion page for the deleted "Songs about Masturbation" Wikipedia page.

due to disagreement about the balance between openness and privacy is unsatisfactory because the two policies are in broad agreement. One way to explain the disparity may be by paying attention to Wikipedia's commitment to a principle of openness and the role this plays in justifying the entire project.

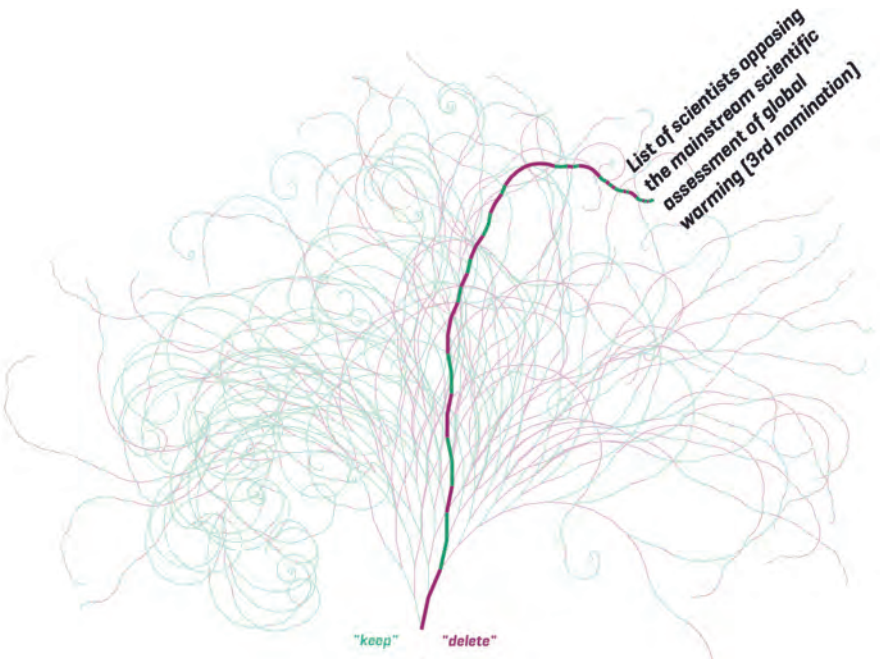
Even if it has its own version of the right to be forgotten, Wikipedia's procedure for "forgetting" is very much its own. Every edit is logged, stored, and debated with reference to the community policies and principles before being approved. Every point of every debate over every edit is also logged, along with the references to the relevant policies. One can therefore find a comprehensive, indelible memory of everything that was ever forgotten, why it was forgotten, who advocated for it, and who objected.

Far from being fundamentally at odds with the idea of forgetting—of closing down material that infringes on individual privacy—the open encyclopedia embraces it. But it manages to reconcile the apparent conflict between open and closed by being open about being closed. This suggests a general strategy by which those working within the open paradigm can feel comfortable within its limitations. If the participation, the policies, the processes, and the end product are all "open," then maybe forgetting need not be seen as an ideological compromise.

The difference between censorship and mere editing is therefore grounded in the community's ability to square its founding principle of openness with some of the new normative considerations it faces. What looks like a substantive conflict between open and closed, public and private, transparency and privacy is dissolved by appeal to a second-order principle of openness, which preserves ideological consistency and editorial sovereignty.

Indeed, publishing indexes of the de-indexed is just one way that the administrative systems and bureaucracy that lie behind Wikipedia's topic pages are subjected to a kind of radical openness. "Talk" pages, where the site's editors deliberate over their activity, have grown faster and are busier than the articles themselves. The "Department Directory" page unveils a bewildering array of governing committees and policy-making processes, from abuse response and counter-vandalism, volunteer recruitment, dispute resolution, and deletion. Every contribution takes place in publicly accessible forums, recorded for posterity in a vast archive of editorial ephemera.[2] Compare this approach with that of traditionally "closed"—or at least, less open—institutions of government, business, or science. Detailed records of internal activity, if they even exist, are usually hidden by default. Even if the official output—a white paper here, a scientific publication there—is made open, the process behind it is not.

2 In this sense, Wikipedia may be reminiscent of the Cairo Genizah, described elsewhere in this volume by Benjamin Outhwaite. The accumulated background pages of Wikipedia are rather like the "ephemera" of daily Egyptian Jewish culture, "piling up in a stratified manner" as a result of the "the rabbinic prohibition against destroying holy writ."



THE KEPT: The 100 longest **Article for Deletion** discussions on Wikipedia which did **not** result in deletion of the article (i.e. it was kept, merged, redirected).
FROM HTTP://NOTABILIA.NET/

The project's commitment to making its inner bureaucracy open and archived is not just an ideological fetish collectively imposed by its community, but perhaps also fundamental to the encyclopedic project. The vast archive of publicly recorded activity serves an important function regarding the encyclopedia's primary content. Wikipedia's aim is to amass the "sum total of human knowledge." This doesn't mean including everything that anyone has or could ever say—it is not Borges' "Library of Babel"—as we can see from the record of deletion and the community's numerous editorial principles. The project legitimizes leaving certain content out by being open about the means and justification for exclusion.

Commercial general encyclopedias never had to justify openly what they'd left out and why (thereby generating significant work for historians interested in their selection criteria). By contrast, Wikipedia's archive of talk pages exists as a record of what was left out and why. The project navigates the contested space between what is considered "the world's knowledge" and what is private, sensitive, irrelevant, unimportant, spurious, or sensationalist. The demarcation of these categories is inherently contestable. By facilitating and archiving such contests openly, the project aims to justify its ambitious claims to totality. A total archive of editorial activity is therefore central to the project's mission to amass the "sum total of human knowledge." ■

REUBEN BINNS *is a postdoctoral research fellow in Computer Science at the University of Oxford, interested in philosophical, technical and legal aspects of personal data, privacy, and the web.*

Wikipedia's aim is to amass the "sum total of human knowledge." This doesn't mean including everything that anyone has or could ever say...

BIBLIOGRAPHY

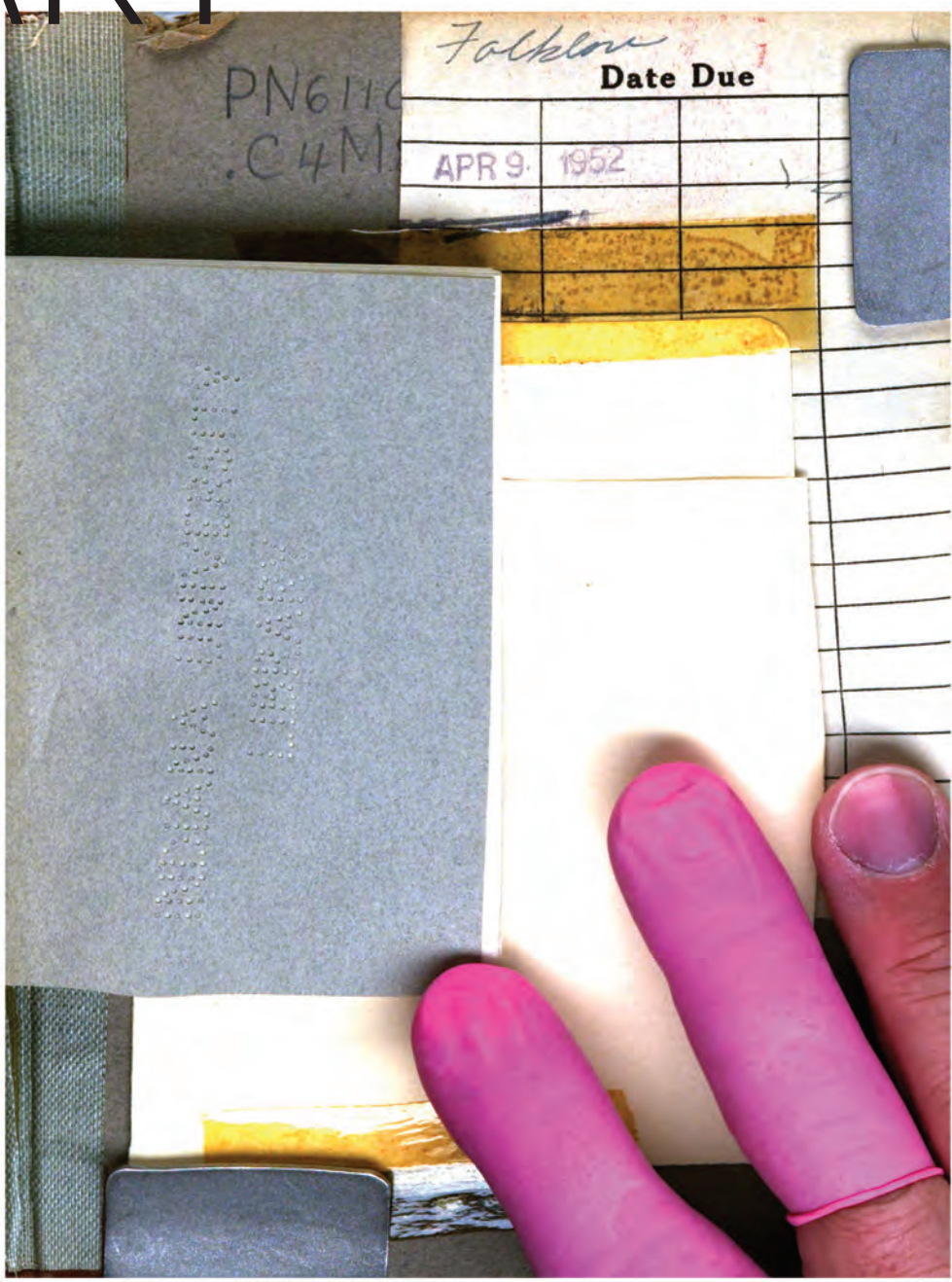
McIntosh, Neil. 2015. "List of BBC Web Pages Which Have Been Removed from Google's Search Results." BBC.co.uk, June 15. http://www.bbc.co.uk/blogs/internet/entries/1d765aa8-600b-4f32-b110-d02fbf7fd379

Wikimedia Foundation. 2015. "Notices received from search engines." https://wikimediafoundation.org/wiki/Notices_received_from_search_engines#English_WikipediaWikimedia Foundation. 2014 "Wikipedia Pages Censored in European Search Results" http://blog.wikimedia.org/2014/08/06/wikipedia-pages-censored-in-european-search-results/

Wikipedia, 2015 "Biographies of Living Persons" https://en.wikipedia.org/wiki/Wikipedia:Biographies_of_living_persons

UNPACKING GOOGLE'S LIBRARY

Google wanted to digitize all the world's books but eventually abandoned that goal. **Mary Murrell** explores the rise and fall of one utopian library project and the emergence of new ones in its wake.



ANDREW NORMAN WILSON, MOTHER GOOSE'S MELODY–164. FROM HIS 2012 SCANOPS EXHIBITION.

IN DECEMBER 2004 GOOGLE REVEALED its Library Project, a hugely ambitious plan to digitize "all books in all languages" through partnerships with some of the largest research libraries in the world—Harvard, Stanford, Oxford, the New York Public Library, and the University of Michigan, to begin—and to make those books accessible online. The news astonished interested observers, eliciting both fear and excitement. Enthusiasts found in it an intoxicating combination of the humanistic and the technoscientific: a new and improved Library of Alexandria, a generation's moonshot, a humanistic complement to the Human Genome Project.[1] The mass digitization of library collections promised to give a future to the past currently "imprisoned" in print form. In so doing it would also give new life to research libraries, open up new lines of scholarly inquiry and practice, and vastly expand people's access to library holdings. Opponents embraced these possibilities, too, but they also feared Google's motives in the project, its will to power, and its increasing control over access to knowledge. Others accused the company of undermining the central tenets of copyright. Brought together in a decade-long saga, these and other contentions around the Library Project swelled into what might be considered an archival fever: one ambitious total archive ramifying into new ones.

Andrew Norman Wilson, The Inland Printer — 164. From his 2012 ScanOps exhibition.

In crucial respects, the Library Project has been remarkably successful. The company has scanned, page by page, more than 25 million books in more than 400 languages.[2] Although it is hard to know for certain how many books exist to be digitized, 25 to 30 million certainly represent a significant percentage. (As a point of comparison, its closest competitor, the Open Content Alliance/Internet Archive has digitized roughly 2 million books.) The Library Project has also survived legal challenges against it. In 2005, authors and publishers sued Google alleging copyright infringement (*Authors Guild et al. v. Google*), and in 2011 the Authors Guild sued Google's library partners over their possession and use of Google's scans (*Authors Guild v. HathiTrust*). In both cases, judges found scanning the entirety of an in-copyright book, for circumscribed uses, to be a fair use under U.S. law—in both lower courts and on appeal.

And yet, despite these successes, Google has quietly forsaken its Library Project, despite being far short of its original outsized goal ("all books in all languages") as well as its pledge to digitize the entirety of the University of Michigan's libraries, its principal partner. After the proposed settlement to *Authors Guild et al. v. Google* was rejected in March 2011, its commitment tapered off significantly. The settlement would have set aside legal differences between copyright owners and Google by opening the Library Project up to extensive commercialization (see Samuelson 2011). Without that potential for revenue generation, the costly project appears to have been deemed too dear even for deep-pocketed Google: scanning capacity was drastically cut in 2011; the Google Books blog was discontinued in 2012; its Twitter feed went silent in 2013; and its staff left or was reassigned. Although the company continues to scan books from libraries, according to partner libraries, it stopped scanning in-copyright books back in 2011, limiting itself now to books in the public domain. This about-face returns to the state of affairs circa 2004, when the announcement of Google's project made such a splash as a bold move forward. By 2011, the Web had changed too. It was no longer in need of high-quality content as it had been in the early 2000s when mass digitization seemed worth the company's investment (Edwards 2011). The moonshot, in short, fell back to earth.

Based on these developments, it is not unreasonable to wonder whether the company might allow its books platform to languish in light of shifting priorities (see Biao 2015; cf. Lemov, this issue). Nonetheless, the momentum around mass digitization has shifted to successor projects such as, in the U.S., the Hathi Trust and the Digital Public Library of America (DPLA)—both of which grew out of the Library Project.

The Hathi Trust began in 2008 as a collaboration among research libraries to pool the digitized books that Google provided as part of their contractual arrangements. It has since grown to include books from other digitizers such as the Internet Archive and from libraries' own scanning initiatives, but its core remains the Google-digitized books. Like a traditional research library or archive (and unlike Google), the Trust's mission is to steward "the cultural record long into the future," with all that that entails (HathiTrust n.d.; see also Christenson 2011). But like Google, it too pursues a totality—a different totality. The Hathi Trust's specific operative aspiration is not the scholar's dream of a "universal library" but rather the technologist's dream of effecting a crucial tipping point, from a print-dominated intellectual infrastructure to an electronic one. By creating one total archive of all books held by its network of research libraries ("curation at scale"), libraries can identify and eliminate the redundancies between their collections, drastically reduce their print holdings, and thus cut out the costs associated with maintaining large and underused print collections (Wilkin 2015). Solving "the print problem" will enable the reallocation of scarce resources to new areas of library activity: institutional repositories, publishing initiatives, redesigned library spaces, data curation, digital preservation, and so on. This explains University of

1 For two examples of the many comparisons to the HGP, see Vaidhyanathan 2012 and Michel et al. 2010. On the HGP as an archive, see also Reardon p. 72 in this issue.

2 The number is now no doubt considerably higher.

Michigan librarian John Wilkin's declaration that December 14, 2004 (the day that Google announced the Library Project) was "the day the world changes" (Associated Press 2004). At last, libraries could look forward to moving beyond the immense burden of their print collections. The problem now, of course, is that Google did not complete the digitization, and it is unclear who will.

Whereas the Hathi Trust emerged through direct collaboration with Google, the DPLA developed in direct critical reaction to the Library Project and, in particular, Google's failed attempt to settle its differences with the publishers and authors. One of the leading opponents of that settlement was Harvard University Librarian Robert Darnton, whose experience working with Google had convinced him that the company's interests were antithetical to those of libraries and the "public interest" (Darnton 2009). In the course of seeking the settlement's rejection, Darnton proposed an alternative "national digital library" which later became articulated as "an open, distributed network of comprehensive online resources" (DPLA n.d.). Officially launched in 2013 with start-up funding from philanthropies and government agencies, the DPLA is not a collection—it has no holdings—but rather a platform that connects dispersed library collections (including the Hathi Trust's). It is more diverse in intention than the Hathi Trust, involving a wider range of organizations (not just elite university libraries) and more diverse types of content (not just books), but it is also more ambitious. Like the Hathi Trust, the DPLA aspires to be yet a different "total archive"—one with a more spatial inflection. In language strongly evocative of early twentieth-century utopians and visionaries, such as H. G. Wells, Paul Otlet, and Robert C. Binkley, who were convinced that microfilm technologies would enable superior scholarly infrastructures, the DPLA seeks ultimately to become a "worldwide network that will bring nearly all the holdings of all libraries and museums within the range of nearly everyone on the globe" (Darnton 2013). To this end, its technical infrastructure was designed to interoperate with Europeana, the European Union-funded Web portal that launched in 2008—and which was yet another response to Google's Library Project.[3]

When, in the early 2000s, libraries forged their awkward partnership with Google over the problem of the printed book, they had sought to manage and to "rationalize" print accumulations. Those attempts, at least so far, appear not to have eased a burden but only to have ramified their responsibilities and increased their accumulations. Library book accumulations have proven themselves to be, more than ever, part of that maddening "universe of things that cannot be disposed of and that keep spawning new things" (Povinelli 2011). Google's Library Project now seems, oddly, if not small at least small*er*. Its successor projects appear to carry more capacious hopes, more intractable obligations—most of which seem remarkably out of proportion to what library leaders understand to be an "era of constrained circumstances" (Wilkin 2015). ■

MARY MURRELL *is currently an honorary fellow in the Department of Anthropology at the University of Wisconsin, Madison. She is writing a book entitled* The Open Book: An Anthropologist in the Digital Library.

3 It is worth noting that, as a project born in response to Google's Library Project, the DPLA, for the time being, involves no book scanning. That job falls to a patchwork of enterprises with varying modes of access: underground "shadow" libraries (see Bodo p. 19 in this issue), some national libraries (though notably not the Library of Congress), the Internet Archive's scanning, boutique library projects, in addition to Google's continued public domain scanning.

BIBLIOGRAPHY

Associated Press. 2004. "Google to Scan Books from Major Libraries." MSNBC.com, December 14. http://www.nbcnews.com/id/6709342#.Vlt3Qd-rSV4

Biao, Andy. 2015. "Never Trust a Corporation to do a Library's Job." *Medium*, January 28. https://medium.com/message/never-trust-a-corporation-to-do-a-librarys-job-f58db4673351

Christenson, Heather. 2011. "HathiTrust." *Library Resources & Technical Services*, 55(2):93–102.

Darnton, Robert. 2013. "The National Digital Public Library Is Launched!" *New York Review of Books*, April 25, http://www.nybooks.com/articles/2013/04/25/national-digital-public-library-launched/

———. 2009. *The Case for Books: Past, Present, and Future*. New York: PublicAffairs.

Digital Public Library of America (DPLA). n.d. "History." http://dp.la/info/about/history/

Edwards, Douglas. 2011. *I'm Feeling Lucky: The Confessions of Google Employee Number 59*. New York: Houghton Mifflin Harcourt.

HathiTrust. n.d. "Welcome to the Shared Digital Future." https://www.hathitrust.org/about

Michel, Jean-Baptiste, Yuan Kui Shen, Aviva Presser Aiden, Adrian Veres, Matthew K. Gray, The Google Books Team, Joseph P. Pickett, Dale Hoiberg, Dan Clancy, Peter Norvig, Jon Orwant, Steven Pinker, Martin A. Nowak, and Erez Lieberman Aiden. 2010. "Quantitative Analysis of Culture Using Millions of Digitized Books." Science 331(6014):176–182.

Povinelli, Elizabeth. 2011. "The Woman on the Other Side of the Wall: Archiving the Otherwise in Postcolonial Digital Archives." *differences* 22(1):146–171.

Samuelson, Pamela. 2011. "Why the Google Book Settlement Failed—and What Comes Next," 54 Comm. ACM 11, 29–31. November.

Vaidhyanathan, Siva. 2012. *The Googlization of Everything (And Why We Should Worry)*. Berkeley: University of California Press.

Wilkin, John. 2015. "Meanings of the Library Today," in *The Meaning of the Library: A Cultural History* ed. Alice Crawford, 236–53. Princeton, NJ: Princeton University Press.

IN THE NAME OF HUMANITY

The total archive is already here. **Balázs Bodó** finds it hidden in the shadows and run by pirates.

AS I WRITE THIS IN AUGUST 2015, WE ARE IN THE MIDDLE OF ONE of the worst refugee crises in modern Western history. The European response to the carnage beyond its borders is as diverse as the continent itself: as an ironic contrast to the newly built barbed-wire fences protecting the borders of Fortress Europe from Middle Eastern refugees, the British Museum (and probably other museums) are launching projects to "protect antiquities taken from conflict zones" (BBC News 2015). We don't quite know how the conflict artifacts end up in the custody of the participating museums. It may be that asylum seekers carry such antiquities on their bodies, and place them on the steps of the British Museum as soon as they emerge alive on the British side of the Eurotunnel. But it is more likely that Western heritage institutions, if not playing Indiana Jones in North Africa, Iraq, and Syria, are probably smuggling objects out of war-zones and buying looted artifacts from the international gray/black antiquities market to save at least some of them from disappearing in the fortified vaults of wealthy private buyers (Shabi 2015). Apparently, there seems to be some consensus that artifacts, thought to be part of the common cultural heritage of humanity, cannot be left in the hands of those collectives who own/control them, especially if they try to destroy them or sell them off to the highest bidder.

The exact limits of expropriating valuables in the name of humanity are heavily contested. Take, for example, another group of self-appointed protectors of culture, also collecting and safeguarding, in the name of humanity, valuable items circulating in the cultural gray/black markets. For the last decade Russian scientists, amateur librarians, and volunteers have been collecting millions of copyrighted scientific monographs and hundreds of millions of scientific articles in piratical shadow libraries and

Pirate librarians regard their libraries as a legitimate form of resistance against the commercialization of public resources...

making them freely available to anyone and everyone, without any charge or limitation whatsoever (Bodó 2014b; Cabanac 2015; Liang 2012). These pirate archivists think that despite being copyrighted and locked behind paywalls, scholarly texts belong to humanity as a whole, and seek to ensure that every single one of us has unlimited and unrestricted access to them.

The support for a freely accessible scholarly knowledge commons takes many different forms. A growing number of academics publish in open access journals, and offer their own scholarship via self-archiving. But as the data suggest (Bodó 2014a), there are also hundreds of thousands of people who use pirate libraries on a regular basis. There are many who participate in courtesy-based academic self-help networks that provide ad hoc access to paywalled scholarly papers (Cabanac 2015).[1] But a few people believe that scholarly knowledge could and should be liberated from proprietary databases, even by force, if that is what it takes. There are probably no more than a few thousand individuals who occasionally donate a few bucks to cover the operating costs of piratical services or share their private digital collections with the world. And the number of pirate librarians, who devote most of their time and energy to operate highly risky illicit services, is probably no more than a few dozen. Many of them are Russian, and many of the biggest pirate libraries were born and/or operate from the Russian segment of the Internet.

The development of a stable pirate library, with an infrastructure that enables the systematic growth and development of a permanent collection, requires an environment where the stakes of access are sufficiently high, and the risks of action are sufficiently low. Russia certainly qualifies in both of these domains. However, these are not the only reasons why so many pirate librarians are Russian. The Russian scholars behind the pirate libraries are familiar with the crippling consequences of not having access to fundamental texts in science, either for political or for purely economic reasons. The Soviet intelligentsia had decades of experience in bypassing censors, creating samizdat content distribution networks to deal with the lack of access to legal distribution channels, and running gray and black markets to survive in a shortage economy (Bodó 2014b). Their skills and attitudes found their way to the next generation, who now runs some of the most influential pirate libraries. In a culture, where the know-how of how to resist information monopolies is part of the collective memory, the Internet becomes the latest in a long series of tools that clandestine information networks use to build alternative publics through the illegal sharing of outlawed texts.

In that sense, the pirate library is a utopian project and something more. Pirate librarians regard their libraries as a legitimate form of resistance against the commercialization of public resources, the (second) enclosure (Boyle 2003) of the public domain. Those handful who decide to publicly defend their actions, speak in the same voice, and tell very similar stories. Aaron Swartz was an American hacker willing to break both laws and locks in his quest for free access. In his 2008 "Guerilla Open Access Manifesto" (Swartz 2008), he forcefully argued for the unilateral liberation of scholarly knowledge from behind paywalls to provide universal access to a common human heritage. A few years later he tried to put his ideas into action by downloading millions of journal articles from the JSTOR database without authorization. Alexandra Elbakyan is a 27-year-old neurotechnology researcher from Kazakhstan and the founder of Sci-hub, a piratical collection of tens of millions of journal articles that provides unauthorized access to paywalled articles to anyone without an institutional subscription. In a letter to the judge presiding over a court case against her and her pirate library, she explained her motives, pointing out the lack of access to journal articles.[2] Elbakyan also believes that the inherent injustices encoded in current system of scholarly publishing, which denies access to everyone who is not willing/able to pay, and simultaneously denies payment to most of the authors (Mars and Medak 2015), are enough reason to disregard the fundamental IP framework that enables those injustices in the first place. Other shadow librarians expand the basic access/injustice arguments into a wider critique of the neoliberal political-economic

NOTES

1 On such fora, one can ask for and receive otherwise out-of-reach publications through various reddit groups such as r/Scholar (https://www.reddit.com/r/Scholar) and using certain Twitter hashtags like #icanhazpdf or #pdftribute.

2 Elsevier Inc. et al v. Sci-Hub et al, New York Southern District Court, Case No. 1:15-cv-04282-RWS

3 While we do not know what his aim was with the article dump, the prosecution thought his Manifesto contained the motives for his act.

4 See United States of America v. Aaron Swartz, United States District Court for the District of Massachusetts, Case No. 1:11-cr-10260

5 Case 1:15-cv-04282-RWS Document 50 Filed 09/15/15, available at https://www.unitedstatescourts.org/federal/nysd/442951/

6 I of course stole this line from Stewart Brand (1968), the editor of the Whole Earth catalog, who, in return, claims to have been stolen it from the British anthropologist Edmund Leach. See http://www.wholeearth.com/issue/1010/article/195/we.are.as.gods for the details.

BIBLIOGRAPHY

BBC News. "British Museum 'Guarding' Object Looted from Syria. *BBC News*, June 5. http://www.bbc.com/news/entertainment-arts-33020199

Bodó, B. 2015. "Libraries in the Post-Scarcity Era." In *Copyrighting Creativity*, edited by H. Porsdam (pp. 75–92). Aldershot, UK: Ashgate.

———. 2014a. "In the Shadow of the Gigapedia: The Analysis of Supply and Demand for the Biggest Pirate Library on Earth." In *Shadow Libraries*, edited by J. Karaganis (forthcoming). New York: American Assembly.http://ssrn.com/abstract=2616633.

———. 2014b. "A Short History of the Russian Digital Shadow Libraries." In *Shadow Libraries*, edited by J. Karaganis (forthcoming). New York: American Assembly. http://ssrn.com/abstract=2616631

Both the curators of the British Museum and the pirate librarians claim to save the common heritage of humanity...

system that aims to commodify and appropriate everything that is perceived to have value (Fuller 2011; Interview with Dusan Barok 2013; Sollfrank 2013).

Whatever prompts them to act, pirate librarians firmly believe that the fruits of human thought and scientific research belong to the whole of humanity. Pirates have the opportunity, the motivation, the tools, the know-how, and the courage to create radical techno-social alternatives. So they resist the status quo by collecting and "guarding" scholarly knowledge in libraries that are freely accessible to all.

Both the curators of the British Museum and the pirate librarians claim to save the common heritage of humanity, but any similarities end there. Pirate libraries have no buildings or addresses, they have no formal boards or employees, they have no budgets to speak of, and the resources at their disposal are infinitesimal. Unlike the British Museum or libraries from the previous eras, pirate libraries were born out of lack and despair. Their fugitive status prevents them from taking the traditional paths of institutionalization. They are nomadic and distributed by design; they are *ad hoc* and tactical, pseudonymous and conspiratory, relying on resources reduced to the absolute minimum so they can survive under extremely hostile circumstances.

Traditional collections of knowledge and artifacts, in their repurposed or purpose-built palaces, are both the products and the embodiments of the wealth and power that created them. Pirate libraries don't have all the symbols of transubstantiated might, the buildings, or all the marble, but as institutions, they are as powerful as their more established counterparts. Unlike the latter, whose claim to power was the fact of ownership and the control over access and interpretation, pirates' power is rooted in the opposite: in their ability to make ownership irrelevant, access universal, and interpretation democratic.

This is the paradox of the total piratical archive: they collect enormous wealth, but they do not own or control any of it. As an insurance policy against copyright enforcement, they have already given everything away: they release their source code, their databases, and their catalogs; they put up the metadata and the digitalized files on file-sharing networks. They realize that exclusive ownership/control over any aspects of the library could be a point of failure, so in the best traditions of archiving, they make sure everything is duplicated and redundant, and that many of the copies are under completely independent control. If we disregard for a moment the blatantly illegal nature of these collections, this systematic detachment from the concept of ownership and control is the most radical development in the way we think about building and maintaining collections (Bodó 2015).

Because pirate libraries don't own anything, they have nothing to lose. Pirate librarians, on the other hand, are putting everything they have on the line. Speaking truth to power has a potentially devastating price. Swartz was caught when he broke into an MIT storeroom to download the articles in the JSTOR database.[3] Facing a 35-year prison sentence and $1 million in fines, he committed suicide.[4] By explaining her motives in a recent court filing,[5] Elbakyan admitted responsibility and probably sealed her own legal and financial fate. But her library is probably safe. In the wake of this lawsuit, pirate libraries are busy securing themselves: pirates are shutting down servers whose domain names were confiscated and archiving databases, again and again, spreading the illicit collections through the underground networks while setting up new servers. It may be easy to destroy individual collections, but nothing in history has been able to destroy the idea of the universal library, open for all.

For the better part of that history, the idea was simply impossible. Today it is simply illegal. But in an era when books are everywhere, the total archive is already here. Distributed among millions of hard drives, it already is a *de facto* common heritage. We are as gods, and might as well get good at it.[6] ■

BODÓ BALÁZS *PhD (1975), is an economist and piracy researcher at the Institute for Information Law (IViR) at the University of Amsterdam.*

Boyle, J. 2003. "The Second Enclosure Movement and the Construction of the Public Domain." *Law and Contemporary Problems* 66:33–42. http://dx.doi.org/10.2139/ssrn.470983

Brand, S. 1968. *Whole Earth Catalog*, Menlo Park, California: Portola Institute.

Cabanac, G. 2015. "Bibliogifts in LibGen? A Study of a Text-Sharing Platform Driven by Biblioleaks and Crowdsourcing." *Journal of the Association for Information Science and Technology*, Online First, 27 March 2015.

Fuller, M. 2011. "In the Paradise of Too Many Books: An Interview with Sean Dockray." *Metamute*. http://www.metamute.org/editorial/articles/paradise-too-many-books-interview-sean-dockray

Interview with Dusan Barok. 2013. *Neural* 10–11.

Liang, L. 2012. "Shadow Libraries." *e-flux*. http://www.e-flux.com/journal/shadow-libraries/

Mars, M., and Medak, T. 2015. "The System of a Takedown: Control and De-commodification in the Circuits of Academic Publishing." Unpublished manuscript.

Shabi, R. 2015. "Looted in Syria–and Sold in London: The British Antiques Shops Dealing in Artefacts Smuggled by ISIS." *The Guardian*, July 3. http://www.theguardian.com/world/2015/jul/03/antiquities-looted-by-isis-end-up-in-london-shops

Sollfrank, C. 2013. "Giving What You Don't Have: Interviews with Sean Dockray and Dmytri Kleiner." *Culture Machine* 14:1–3.

Swartz, A. 2008. "Guerilla Open Access Manifesto." https://archive.org/stream/GuerillaOpenAccessManifesto/Goamjuly2008_djvu.txt

UNENDING ARCHIVES

Aleph or Library? Work from the **Artist Collective UA** explores whether art can be an archive, or an archive art.

An artwork might intuitively seem like an object more likely to be found *in* an archive than to constitute an archive in its own right. But if we recall Jorge Luis Borges' Aleph – the object or "point in space" that opens out to reveal a whole universe – we can begin to see how a work of art can behave like an archive. This is less abstract than it might sound. Aby Warburg, for example, was inspired by the idea of artworks as partial cultural archives that could be wrenched from their historical eras and reassembled to form a new way of visual knowing. He called the resulting panels *Mnemosyne Atlas*, though the project remained unfinished at his death.

Alternatively, an artwork can be built from a small but complete informatic system, at which point it forms the limiting case of its own world:

And, as can be seen elsewhere in this issue, archives can also form the material for more ambitious (and playful) classificatory experiments (Hess, de Croehling, Prévieux).

Always, however, the archival function of the artwork – that is, to be a transparent window on to external events – is incomplete. Archival art always presents us with a figure/ground confusion: do we see *through* the work or look *at* it? What I want to propose is that in archival artworks we have a constructive counterpart to the de(con)structive postmodern "crisis" of representation. Rather than systematically breaking down the stability of the subject, the meaning of the work, the mortality of the author, archival work presents a *positive* instability. Such instability was described recently by N. Katherine Hayles in her talk A Theory of the Total Archive:

> *A total archive is of course literally impossible, but in imaginative literature, there are two ways to achieve it, as Borges has taught us: infinite expansion ("The Library of Babel"), and infinite compression ("The Aleph"). Apparent opposites, the two cycle continuously into and through one another, as do outside/inside in a Möbius strip or interior/exterior in a Klein bottle.*

To Borges' Aleph, Hayles adds his Library – a vast arrangement of all knowledge (and nonsense, for that matter). Both the Aleph and the Library are paradoxical: because the Aleph condenses space, it must expand; because the Library is complete, it has an arbitrary and limiting structure. Hence Borges' story ends with the idea of an "eternal traveler" crossing the library and eventually finding that "the same volumes were repeated in the same disorder (which, thus repeated, would be an order: the Order)."

FIG. 1: THE ARTWORK AS DOCUMENT. Panel from Aby Warburg's *Mnemosyne Atlas,* an unfinished, thematic and aggressively anachronistic art historical project, now reconstructed in a fascinating website hosted by Cornell University.

FIG. 2: THE ARTWORK AS INFORMATION. Laszlo Moholy-Nagy, *Constructions in Enamel I, II and III,* 1923. Identical works at different scales, purportedly made by an enamel factory which had received the specifications over the telephone.

FIG. 3: THE ARTWORK AS REPOSITORY. Fabienne Hess, *Hits and Misses* (Talbot Rice Gallery, Edinburgh 2015). The work is a fabric showing every available image from the University of Edinburgh"s digital archive.

Likewise, where the artwork expands most fully it also becomes the most constrained, resistant to explanations that break apart the contradictions in the work. Relics are like this: they are either legitimate, sacred, non-made instantiations of the past alive in the present, or they are artificial, profane objects that tell stories of their non-miraculous provenance. We want them to be both, but they can't be. Forgeries have this quality too. They are overburdened with historical meaning and therefore appear to have none.

Hayles calls the system that constrains a "total archive" an apparatus of control – again this sounds abstract but she is referring to quite concrete administrative systems: governments, corporations, universities. Interpretative inflexibility tips over into institutional control. So archival artworks can build worlds, and they can also wield authority, demanding much more than simply to be looked at or considered. We are on the outside, and they invite us in.

THE LONDON-BASED ARTIST COLLECTIVE "UA" repeats a "Universal Hymn" at each of the group's meetings. The hymn was "composed" by electronic music pioneer Peter Zinovieff, who gave the group the instruction to say/sing the letter A in any pitch, at any volume, in any pronunciation, for any duration: "Universal Hymn is A." And he might have added: "Universal Hymn is Always A", because, following the logic of ritual, any utterance of it is authentic.

A work like the Universal Hymn is always the same piece, across all performances, but it doesn't rely on any idea of an original which can be copied. The Hymn is a minimal auto-archive, an Aleph. And it opens out into... whatever comes after – Live Action Role Play, collaborative sculpture, performance, song, poetry. So the Hymn is

FIG. 4: Miriam Austin, *Prosthetics for Hostile Contexts* (2015). Part of the exhibition "Groundwork", at the New Art Centre, Roche Court, Salisbury.

FIG. 5: Miriam Austin, *Prosthetics for Hostile Contexts* (detail).

FIG. 6: Anna Hughes, *Ouroboros* (2014), which aims, in Hughes words, for the "hopeful elevation of these objects into ritualistic talismans".

FIG. 7: Anna Hughes, *Looped Among Islands* (detail).

a social technology, like all good ritual.

The opposing archival tendency, the Library, is also present in UA's work. Through installations they build worlds in which rituals have a place. This is explicitly archival, implicitly totalising. Borges' Library contains every possible arrangement of letters – all words, all sentences, all books – and so meaning is randomly distributed can only be found by hard work. As Hal Foster put it in his 2004 essay on the "archival impulse" in art:

> *Although the contents of this art are hardly indiscriminant, they remain indeterminant like the contents of any archive, and often they are presented in this fashion—as so many promissory notes for further elaboration or enigmatic prompts for future scenarios.* (Foster, 2004)

Rebecca Lemov's drawings (p. 30) from the *Database of Dreams* are like this. They are, in the phrase of Michael Taussig's that Lemov quotes, "fragments that are suggestive of a world beyond, a world that does not have to be explicitly recorded and is in fact all the more "complete" because it cannot be completed."

For example UA member Miriam Austin's installations and performances are always already old – pristine, but only because they have been maintained in such good order, practiced with such care. The archival function of these works is precisely that they conjure a totality: some objects are familiar.

But the flowers are not quite natural, glistening but dry – apparently preserved in rubber. The knives have no handles and on inspection look more like the clavicles of an extinct metallic creature. The tables on which objects rest are split down the centre, creating a gap down which substances slide. Precision here is unnerving, imprecise. The absolute authenticity of ritual is grafted onto an unreal world in just the way authors use small "real" details in magical realism. The work is a world through which you can walk, and it opens out endlessly as a space of imagined vigilance, order, meaning.

Another UA member, Anna Hughes, makes work that is superficially similar, but intentionally less comprehensive in the world it creates. Here the exertion of a strange order brings us closer to the minute perfectibility of the Aleph than the grandeur of the Library. If enough force is applied to the arrangement of natural detritus it will cohere into a taxonomic singularity.

This is the straightforwardly historical element of Hayles' idea of infinite compression/expansion: classifications require raw data, but data, as any historian or sociologist will tell you, is always cooked (Gitelman, 2013). This has not been a problem for the architects of archives – far from it. The presence of classification *in* data has motivated the most ambitious of all projects, from the archive of all human happiness (Jardine, p. 48) and of all human movement (Laemmli, p. 48) to the composition of the great encyclopaedias. As Simon Schaffer has pointed out, the Enlightenment encyclopaedists, in recognizing the multitude of possible "projections" on which their maps of knowledge could be based, still worked at "collecting knowledge into the smallest area possible and of placing the philosopher at a vantage point high above this vast labyrinth (Schaffer, forthcoming)." Universal *because* particular, a paradox that again pushes us to situate the artwork in a world, not just of ritual action but of half-reasoned order.

But the encyclopaedia, in one important sense, also eschews all classification: it presents itself by means of an anti-order – the alphabet. UA member Emily Jones comes closest the absolute plane of alphabetic arrangement in her list-poems, twitter feed and installations. But where Borges famously saw surrealism in the encyclopaedia"s juxtapositions, Jones draws out the affect of an unending (if not total) archive. Here is a poem she has written for this issue:

ROSEATE SPOONBILL ARRIVING AT THE COAST (IT'S THE RIGHT MOMENT)

structural impurity
essentially open
standing position
answer yourself
ten heavens
home office
life force

The terms are indexical, in both the paratextual and linguistic sense: insofar as they fail to interrelate, they have to point elsewhere, to some *other* state of affairs, a location in the book of the natural and technological world. This is the assertion of an intuitive order onto an overwhelming slew of information. Jones' intuition, however, is not formal but ecological: the juxtaposed phrases, titles and imagery range over the natural and the personal, the machinic and political. We have the materials for salvage, they seem to say, so let's use them. The final archival meaning of a work is that it can be an "ark", just like the first (and presumably last) museums. The impersonality of Jones' writing is not a hopeless gesture to a world after humans or a non-human ontology, rather it acts to distribute authority equally amongst its component parts (the ark itself, of course, was built to a sacred design). ■

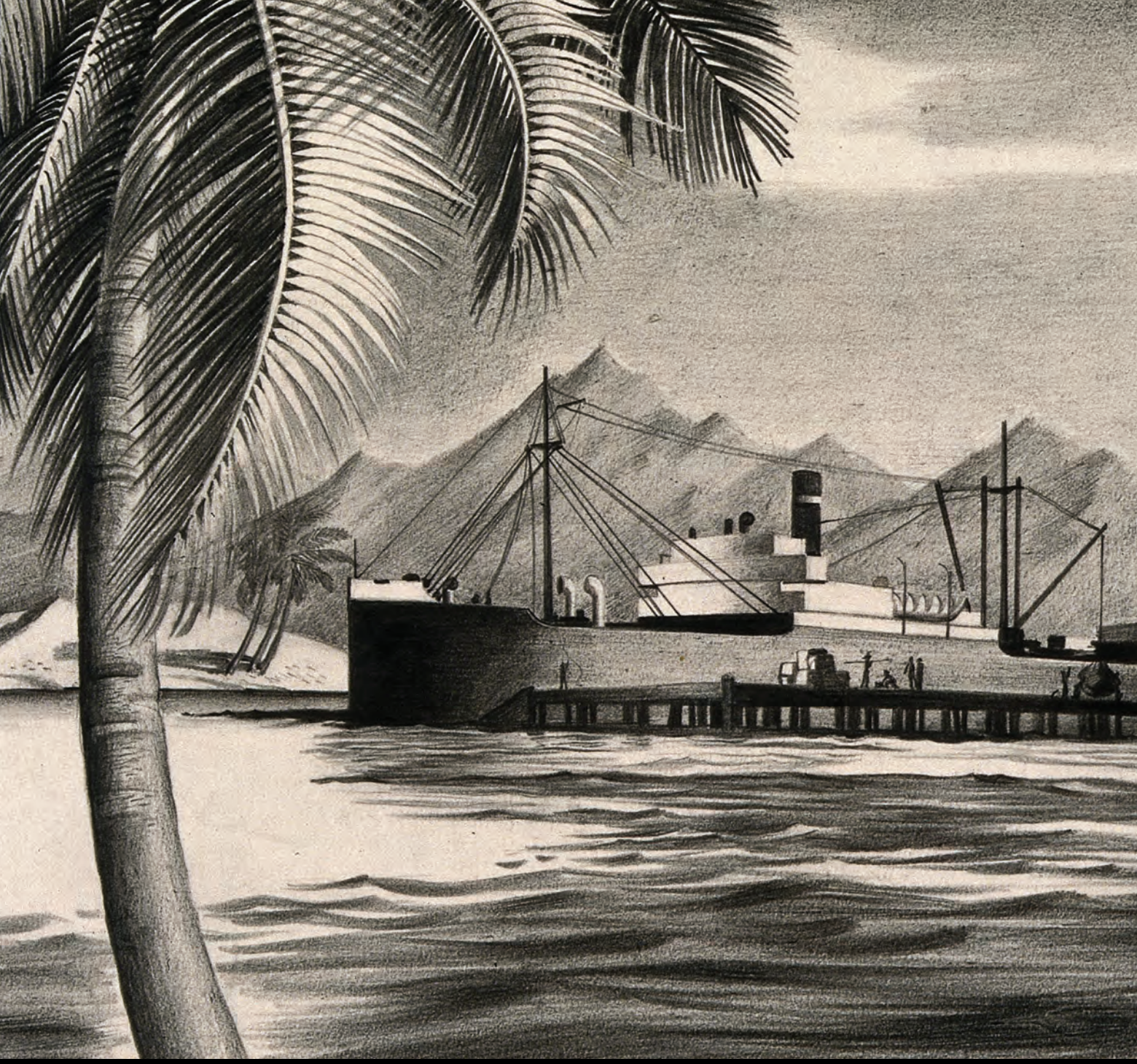
THE UA ARTISTS COLLECTIVE *includes* **MIRIAM AUSTIN, MATTHEW DRAGE, PAUL GWILLIAM, ANNA HUGHES, BORIS JARDINE, EMILY JONES** *and* **RUMI JOSEPHS.** *Their work is featured in the issue 18 of Art Licks.*

BIBLIOGRAPHY

Gitelman, Lisa (ed.). 2013. *Raw Data is an Oxymoron.* Cambridge, MA: MIT Press, 2013.

Foster, Hal. 2004. "The Archival Impulse" *October* 110:3–22

Schaffer, Simon. Forthcoming. "Afterword", *Re-Visioning the World: Myths of Universal Knowledge and the Aesthetics of Global Imaging.*



Fragments *of* plague

Branwyn Poleykett, Nicholas HA Evans and **Lukas Engelmann** are rethinking the role of the visual in the creation of a total archive of the Third Plague Pandemic.

THE THIRD PLAGUE PANDEMIC likely originated in Yunnan, China, in 1855. After arriving in Hong Kong in 1894 it spread from there across the globe. The epidemic was a testing ground for bacteriological medicine and public health, and galvanized the hygienic planning of colonial cities while also testing the rationalities and logics of colonial governance. Riots and resistance characterized some responses to colonial policies, while in other places there emerged complex political entanglements of colonial and subject agency. As members of the "Visual Representations of the Third Plague Pandemic" project, we are building a digital image archive of plague, including photography, scientific illustration, maps, and caricatures from all of the continents touched by plague between 1860 and 1950. This constitutes a new total visual archive of the first pandemic photographed on a global scale.

The visual collection that our project draws together constitutes a new intervention into the existing archival evidence of the pandemic as a global event. This new arrangement queries photography as a transparent historical source. When we encounter images, we enter them into a database that will eventually become an open access resource for historians of science,

4


5

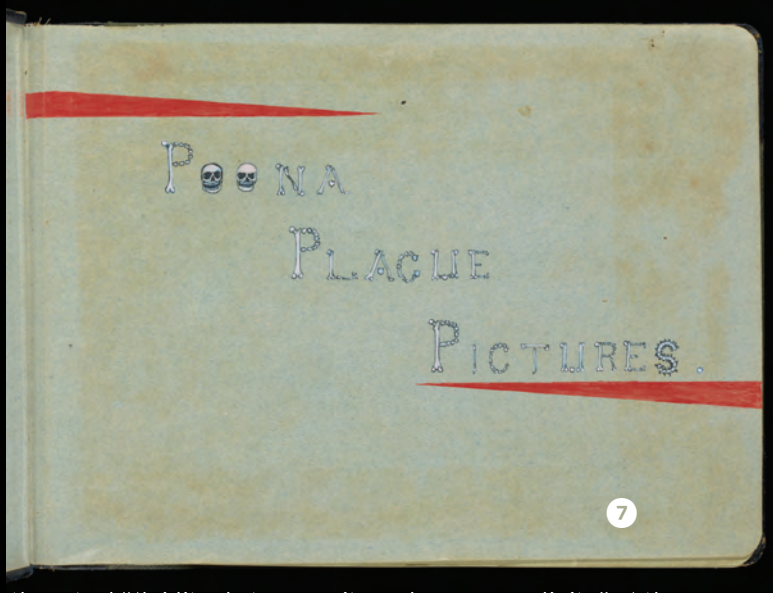
medicine, and visual culture. The database allows us to make lateral connections among San Francisco, Chinatown, rural South Africa, the docks of Glasgow, the *chawls* of Bombay, and the Sino-Russian border. Its agglomerative logic lets us slice across scales and unmake imperial territorializations that are historically taken for granted.

As a total object, or in fragments, our newly created visual archive of plague helps us to see in a different way, but it also raises questions about how to sift and sort these images: What kinds of lateral connections can be made? Which relations require detailed historical textual exposition and contextualization? How do the images collected deform and reframe the meta-category to which they refer? Which transversal, transnational connections are arbitrary, or artifacts of colonial framings, and which might be crucial? Here we reflect on the creation of a total collection of digital surrogates abstracted from colonial archives, and how this process of collecting, creating, and completing historical images disrupts our sense of a historical event. Using technology to re-encounter and curate the past leads us to query what our analysis does to our strange, discontinuous, always-unfinished, total archive of plague.

Photographs are often found lying oblique to the textual archives, boxed up out of the way or cataloged separately. The colonial archive–for example, in British India–was frequently constructed in an effort to comprehensively enumerate the population, yet photographs often have no place in the official archives of government. Instead, photographic collections–found as they are in the private scrapbooks of doctors, missionaries, and individuals at the edges of official power–can speak to the impossibilities of knowing everything. Take, for example, an album of the plague outbreak in the Indian city of Poona, a site of government panic and popular rebellion against interventionist plague measures. Juxtaposed in this single collection are holiday photos of ancient monuments (❹), a quaint image of dining room (❺), and gruesome portraits of dying plague patients (❻), all of which are preceded by a macabre yet strangely humorous title page (❼).

8

6



7

Photographs in their "archival context" frequently challenge the archive's rigid frames of intention and its totalizing ambitions. Yet in much historical work on plague, these self-same photos are put to work in "illustrating" arguments with little regard for the technical, perceptual, and political work that photography can do. Take, for example, a photograph (8) of the plague outbreak in the Chinese quarters of Honolulu, Hawaii, in 1899: we know almost nothing about this scene, but it appears to depict the public washing of a group of naked men, a hygienic process. This photograph has a dramatic appeal that encapsulates a moment of obsessive sanitary policy undergirded by a racialized logic and a threat of violence, which traditional narratives have identified with the supposedly global experience of the Third Plague Pandemic. When it appeared on the front cover of a recent book about the global pandemic (Echenberg, 2007), the picture was therefore called upon to represent the entirety of the pandemic. It thus lost its position as an indexical datum in a vast collection of photographs preserved in the Hawaiian State Archive, a collection initially commissioned by the local Board of Health to document a local and rather minor outbreak of plague in Honolulu.

Yet within its own collection, it is nonetheless an unusual picture. The Hawaiian State Archive preserves 400 photos in this archival box, but the vast majority portray empty houses, streets, and places and produce a visual frame of plague void of any actions, events, or practices and instead concerned with mapping and understanding the epidemic's ecology. Yet this particular photograph, when disconnected from its archival locality and the framings of a state's history, and when freed of the specific sanitary regime under which it was created, has become a fixed representation of a pandemic that can cause complex and contradictory narratives to fold together into a simplified historical picture of plague. By bringing this single "iconic" photograph back into relation with its series and its many histories, we are reminded of the actual "fluidity, heterogeneity and even serendipity" of the archive, in which the making and preservation of photographs is never sufficiently turned into a unified practice (Edwards 2001:4). To (re)place photographs in our archive is one way of reworking these particular pictures' possibilities for relating to the picture of the whole. It allows them to form a new set of relations based upon logics of genre; as such, it allows them to rediscover their own contextual specificity as they begin to relate to other photos through relations of likeness, similitude, mimicry, and resemblance. Our project of total accumulation is bound to absurdity, but as a process, it creates relations that place demands of meaning upon the photo, rupturing its wholly illustrative status. ■

NICHOLAS HA EVANS *is thinking about commissions of inquiry*, **BRANWYN POLEYKETT** *is working on French colonial science*, **LUKAS ENGELMANN** *is examining epidemic mappings; all are currently postdoctoral researchers at CRASSH, University of Cambridge.*

BIBLIOGRAPHY

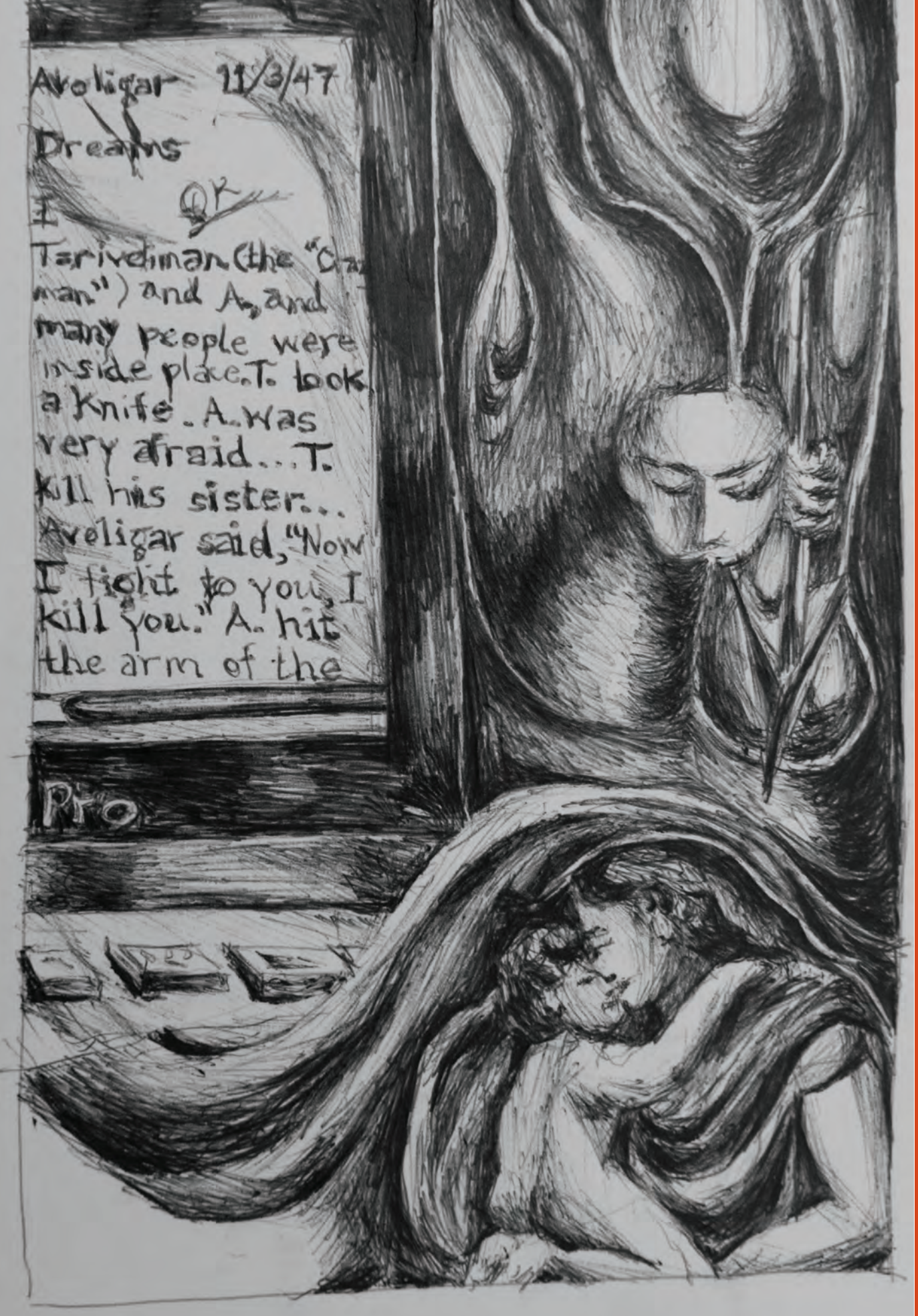
Echenberg, Myron. 2007. *Plague Ports: The Global Urban Impact of Bubonic Plague, 1894–1901* NYU Press

Edwards, Elizabeth. 2001. *Raw Histories: Photographs, Anthropology and Museums*. London: Bloomsbury Academic.

ACKNOWLEDGEMENTS

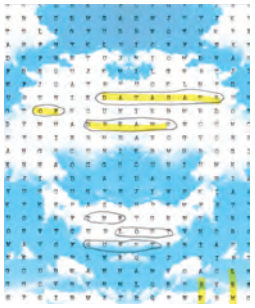
The authors acknowledge the support from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007–2013)/ERC Grant Agreement n.336564

IMAGES USE BY PERMISSION
Figure 1, 2, and 3: Wellcome Library, London. Wellcome Images
Figure 4, 5, 6, and 7: Getty Research Institute
Figure 8: Hawaii State archives.


Aveligar 11/3/47
Dreams
I
Tarivelman (the "C
man") and A, and
many people were
inside place. T. took
a knife. A. was
very afraid... T.
kill his sister...
Aveligar said, "Now
I fight to you, I
kill you." A. hit
the arm of the
Rio

WHAT ESCAPES THE TOTAL ARCHIVE

Rebecca Lemov relates how the stories in the a "database of dreams" leak out of the edges, and sometimes overwhelm totality with particularity.



This work derives from Rebecca Lemov's book *Database of Dreams:Forgotten Archive of How to Be Human* (Yale Univ. Press, 2015)

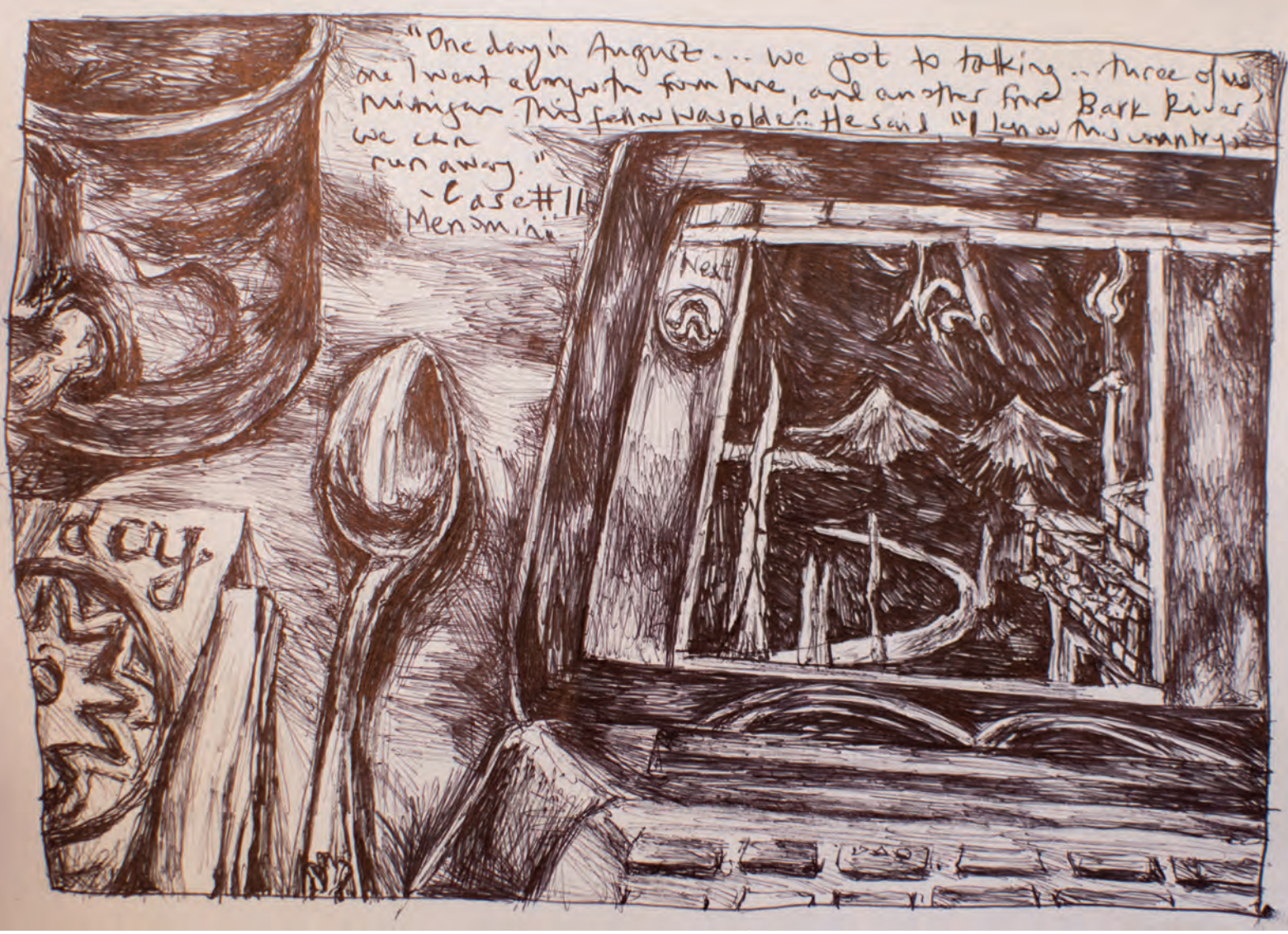
LEFT: Illustration of a dream of an Ifalukan who dreamed about Tareveliman wielding a knife in November 1947; the dream is preserved in the Kaplan archive, which I found at the Library of Congress. DRAWING BY THE AUTHOR.

1.

Pursuing the twentieth-century dream of capturing all sociological data in a single clearinghouse, a group of American social scientists in the mid-1950s attempted a bold, if not completely unprecedented, experiment. They would test the limits not only of content (*what* was collected) but also of format (*how* it was collected, saved, circulated, and distributed). The resulting data set of data sets, which I call the "database of dreams," but which its creators referred to by the somewhat less evocative *Microcard Publications of Primary Records in Culture and Personality,* took shape between 1955 and 1963. Meanwhile, its more extensive vision—the total archive it portended and evoked containing all ephemeral data from the domain of subjectivity collected from peoples around the world, and available in turn across the globe—never did come about. Yet its would-be creators spoke of it as if to invoke it into existence.

At first the group worked collectively on what they saw simply as a pilot project, one that would contain hard-to-capture, ephemeral "human documents" such as life histories, nighttime dreams, daytime wandering thoughts, and psychological test protocols. If this succeeded, and if funding continued apace, they could then address the "all" of social scientific data. So it was that this set of experts, including the project's spur, the young Kansas-based psychologist Bert Kaplan; his mentor, the Philadelphia anthropologist A. I. Hallowell; the Harvard child-development anthropologist John Whiting; the Iowa sociologist Roger Barker; and the micropublishing entrepreneur Webb Thompson, Jr., chose a subset of the "all" and went about preserving what they saw as endangered, ephemeral data, hoping for their wide circulation. Funded by the National Research Council, the National Institutes of Health, and the University of Kansas, among others, the committee built a data pipeline and storage infrastructure. The result held troves of social science materials from generations of fieldworkers who studied people considered faraway, nonliterate, less advanced, or non-Western (even if, as in many cases, they were American Indians), bringing them together in what was then the most advanced technological formats allowing scalability: compression devices for storage (the Microcard) and expansion devices for access (Readex machines). The former were opaque cardboard cards that held compressed texts (each 8 ´ 11-inch sheet shrunk down to the size of a fingernail); the latter were reading devices that magnified the miniaturized texts back to a legible size on a screen.

The project had an "at your fingertips" component: use of portable and desktop reading machines would make it accessible with the greatest ease of movement, an early version of "just a click away" data available on mobile or stationary devices. As a result, hundreds of libraries around the world acquired the prototype of some 20,000 miniature pages, even though it worked awkwardly with the low-cost "pocket reader" (no larger than a king-size packet of cigarettes, it was convenient, but not pleasurable for long-term reading), thus spreading far and wide its globally gathered data. The network meant you could effectively hack into the dream or intimate life details of a person you had never met and likely would never meet, and in fact this contraption made it more likely such a meeting would never occur: it made fieldwork transmissible and thus less necessary for experts of various stripes not inclined to take long trips to the



ABOVE: "I know this country. We can run away..." (Case No. 9, Wisconsin, Microcard Publications of Primary Records in Culture and Personality 1956). DRAWING BY THE AUTHOR.

Gran Chaco or other places to gauge remote psychologies.

The group brought together a concatenation of advanced methods in social science, library science, and related fields. They borrowed interview prods from psychoanalysis, situations from Cold War behavioral science, experimental imperatives from anthropological fieldwork tradition, and microphotographic technologies (described above). Not least, they employed the capabilities of projective tests to tap into the unwitting subject's inner concerns. Considered within the context of the fieldwork relationships that sustained them, these tests became incantatory devices. A man named Tariveliman in the southwestern Pacific became devoted to the anthropologist Melford Spiro, who gathered his dreams and tested him with the Rorschach and other protocols as part of his dissertation research. While he lived on the atoll in 1947–1948, Spiro tested Tariveliman with a "battery" of multiple instruments and found his respondent's answers so deranged that his results did not even fall within the range of testing parameters. Tariveliman in turn followed the anthropologist around each day, longingly picking up his empty Lucky Strike containers and caressing the inkblot cards. While in Spiro's tent one afternoon, Tariveliman reminisced about the Rorschach cards, declaring, "Those pictures you showed me, I liked them very much." Ironically, Tarev's own records did not make it into Spiro's data set as collected in the Microcard archive, "156 Rorschachs, 126 Modified T[hematic] A[pperception] T[ests]s, 83 Stewart E[motional] R[esponse] T[est]s, 82 Bavelas M[oral] I[deological] T[est]s, and 54 Dreams of Ifaluk Men, Women, and Children" (Spiro 1956). He did, however, appear by name in two of the 54 dreams.

The Rube Goldberg functioning of this data storage device was supposed to make it possible for any researcher to access a Burmese or Balinese person's most private thoughts and even her last night's dream; or, as in the case of Tariveliman on the island of Ifaluk, the frightening dreams other islanders had of him. Yet, despite some

ABOVE: "corn fields was our sleeping place..." (Case No. 9, Wisconsin, Microcard Publications of Primary Records in Culture and Personality 1956). DRAWING BY THE AUTHOR.

success, the device did not flourish; that first version was also the last. In a long tradition of imaginary machines, it remained, for some time, difficult to locate. Instead, the archive's chosen format and pre-existing methodological and theoretical frameworks became obsolete, so that the data itself did not so much disappear as lapse into abeyance, latency, irrelevance. (Note that although the original READEX reading machines for opaque micro-formats themselves gradually became obsolete, next-generation analog machines and finally the ST200X Series Digital Film Viewer allowed me to access the data when I began my research around 2007.)

2.

In gathering all this, its creators hoped to capture something like "everydayness." Not the structure of Ifaluk life or the template of Balinese religious rituals, but the actual, nearly irreproducible feel of these things, the what-it-was-like of ordinary life experiences as they went by, something like what Borges called the "irrecoverable colors of the sky." It was an index of social-scientific confidence that they felt finally these nonmaterial phenomena could be materialized.

Here is a series illustrating a life history found in the "database of dreams," told by a middle-aged Menomini Indian man labeled Case No. 9. The longer story appears in the final version of my book (Lemov 2015). A brief summary of his story is that when Case No. 9 was a young boy, around 12, he was diagnosed with tuberculosis and sent to an Indian Sanitarium in Iowa, several states away from his reservation home in the woods of Wisconsin. He did not see his parents for several years, and one day arrived at a plan to run away. With two friends, he snuck out of the asylum, hid in corn fields, and set out on the journey home, encountering along the way both kindness and cruelty from strangers. Finally he experienced a miracle: a hostile farmer with a hatred of Indians began hunting them with a pistol after they asked if they could sleep in

his barn. They hid in fields of tall wheat, listening to the farmer's dog search them out. But even though the dog crossed near them several times, he never betrayed them. "All was still," Case No. 9 recalled. They took an opportunity to run and made their way to safety. When the boy returned to his hometown, no one recognized him at first, for he had grown taller into an adolescent frame. Finally he was reunited with his father; this was the only time in his life the boy saw him cry.

Kaplan's database was a slumbering giant. Rare forms of data rested there. They did not exactly live, and they did not exactly die. This was true of Case No. 9's story of his escape from the asylum, as well as many other stories. The man the boy became had always intended to send it to *True Story* magazine, but had never gotten around to it ("I heard they give prizes. I think of that a lot"), so when the anthropologist appeared to take down his life when he was 44, he readily told the story as if it had already been written, which is exactly how it reads.[1] Like Case No. 9's story, myriad life histories in the databank remained in a curious unresolved state, preserved but not published, archived but not really available. They rested in limbo. That is where I found them half a century later.

Today, professional social scientific archives such as the Harvard "Dataverse"—a repository created to "Share, publish, and archive your data. Find and cite data across all research fields" that now holds 59,287 data sets—function much as Kaplan and his group imagined their database would. Its contents symbolize, according to Gary King, the Harvard political scientist who runs it, the very future of the social sciences, which lies in the realm of intensive data collection. Meanwhile, the other side of Kaplan's twin experiment, the imperative to collect the always-fading-away quality of everydayness, carries on in diverse public-oriented projects such as Storycorps, Radio Diaries, and other neo-documentalist projects that work with recovered archival materials. In the case of Storycorps, pursuing the creed that "every story matters," engineers record day-in-the-life stories that re-evoke the possibility of a global archive. Born from Dave Isay's original StoryCorps booth opened in Grand Central Terminal Station in 2003, it now boasts a mobile van, an

ABOVE: "Finally, we wound up in Chester, Minnesota..." DRAWING BY THE AUTHOR.

1 He concluded matters by telling the anthropologist collecting his life history, "Well, I guess that's all George. There wasn't much to my life. The main incident was the time I run away from that sanatorium. That's the only thing that seems interesting to me." He had never told the story to an outsider before, and perhaps it would never have been told had Spindler not come along.

ever-widening mandate, a set of prestigious awards, a podcast, and a mobile app for collecting. Particular populations such as 9/11 witnesses or LGBTQ people in America are the focus of special story-collecting efforts, but the overall aim is "everyone," as both potential listener and potential taleteller. Case No. 9 might easily have found a home there; in fact, I have been contacted by Radio Diaries to explore this possibility.

3.

Writing a history of this effectively mislaid and definitely neglected archive occupied me for eight years; during one of those years, I made many drawings (including the series shown here), mostly instead of writing. My method was to take an item or a detail from the database of dreams and find a way to represent it on a page by reproducing the words from my pdfs of the archives, which I had saved by means of the ST200X machine. I then surrounded the copied text with images that I borrowed and redrew from certain websites caching 1940s- to 1970s-era scientific and sci-fi images. Finally, I drew all this within the frames—such as my laptop or the kitchen table—by means of which I was seeing these things. In all of these, my medium was a black or blue ballpoint pen and a sketchbook, both of which could easily be transported anywhere I went. I finished quite a few of the drawings, including one that explored a line from dream collector Dorothy Eggan, who wrote of frequently reassuring Hopi Indians in the preface to her data, "I don't want your secrets; just give me your dreams." Then, I went back to writing primarily.

Here I'm highlighting the drawings because they were a way of making sense of what, in the end, was "total" about this erstwhile if not quite vanished effort. It was an archive that, in yearning toward totality, of course failed to realize it. Its authors failed even to make the step beyond their own "pilot" project, and as a result, the prelude became the seemingly final act. My conclusion was that the project was illuminating for this very reason: totality did not lie where they thought it did. It is in the striving for completeness rather than the arrival at completeness that its significance lies; likewise, it was in the dream *of* data and not the data of dreams that the project was important. Reflecting on the place of his drawings in fieldwork notes, Michael Taussig recently observed, "The drawings come across as fragments that are suggestive of a world beyond, a world that does not have to be explicitly recorded and is in fact all the more 'complete' because it cannot be completed" (2011:13). I think that's what these were for.

These drawings represented (to me) my own story as a researcher: my attempt to make sense of what the quest for totality means: the "fantasy of total information." Drawing them brought out for me the slumbering quality of this almost infinitely expansive database, whose fortunes so rapidly reversed. I was trying to draw my interpretation of the world in which these stories now awaken.

Today, the dynamic of the total archive continues to operate. We look for totality in the wrong place. I like to think of the "database of dreams" as an unintended, beautiful modernist masterpiece: an anthropological *Ulysses* with multiple authors, some named, some anonymized, others unidentifiable. Pastiche arose where universality was sought. As with this volume's other examples of mid-twentieth-century total archives that appear baroquely partial in retrospect—Jardine's midcentury happiness metrics, Laemmli's global dance archive, and Kaplan's endangered-language storehouses—the archive imperfectly representing totality touched on something vitally important.

Without necessarily wanting to, such scientific subjects of modern change were forced to engineer their own lives. Their stories fill the database of dreams. They *are* the database, filed along with the accumulated test evidence and accompanying dreams, rendering it a different kind of total archive. Their very incompleteness, the "pathos of the perishable format" to which they succumbed, and the mishmash of the social-scientists' dreams with the subjects' dreams—each also dreaming of the other—form a palimpsest for an ever-growing totality. ■

REBECCA LEMOV *teaches in the History of Science Department at Harvard University*

BIBLIOGRAPHY

Lemov, Rebecca. 2015. *Database of Dreams: The Lost Quest to Catalog Humanity*. New Haven, CT: Yale University Press.

Spiro, Melvin. 1956. Data set. *Microcard Publications of Primary Records in Culture and Personality*, v. 1. LaSalle, WI: Microcard.

Taussig, Michael. 2011. *I Swear I Saw This*. Chicago, IL: University of Chicago Press.

ZEBRAS, BLANKS AND BLOBS

How can we work with vast digital collections? Artist Fabienne Hess explores the content and scale of an online image database.

ALL IMAGES © THE UNIVERSITY OF EDINBURGH

IN EARLY 2015 I ASKED THE UNIVERSITY OF EDINBURGH TO SEND ME every out-of-copyright image from their vast digital archive. With this I made the work for my exhibition *Hits and Misses (from the archive)* for Talbot Rice Gallery in Edinburgh. I came to the 20,000-plus images as an outsider, knowing I was going to be overwhelmed. Roughly a third of the university's digital archive had never been accessed at all; clearly one wasn't meant to look at it in one go. But I wanted to find out what would emerge from this situation.

I spent months looking at the images over and over again. I say "looking," but it was more like wrestling with or trying to tame them. One needs patience and focus to repetitively scroll through thousands of medieval manuscripts without getting exasperated.

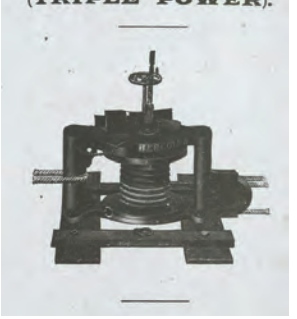
I didn't particularly look for trends, but accidentally stumbled upon groups of images that had similar formal qualities. At first I started to notice the blank pages, carefully photographed, sometimes with just a trace of an image shining through from the other pages. Then the many book covers bearing no information caught my eye, then the marbled endpapers, then triangular shapes, black and white stripes, black blobs, etc. I tagged the images according to these characteristics as I went along and later grouped them. My only constraint with these categories was that they couldn't be a search term the university had already indexed. At some point I could only look at the images with the categories in mind. The more I looked, the more I saw. These categories even transferred into real life; for example, I started noticing zebra textures in people's clothing, or in fences.

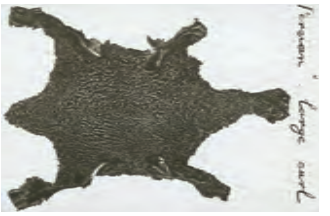
These basic formal characteristics were all I could register under this avalanche of images. Digging into the story behind the images or the objects they depict—how they came to be part of the archive, for example—would have been too much. But what I could do is *see*. Seeing without asking questions, the way we see when drawing in perspective; a seeing that would just register a black square in Malevich's *Black Square* painting. A robotic seeing that in fact some computers are capable of already. The Rijksmuseum allows users to search its online collections by color; so does the British Library. The latter even indexes triangles, curves, and monochromes, but, alas, no black blobs. Incidentally, this was the category that caught my eye last.

Looking at the university's archive reminded me of surfing the internet, where I know there is always more around than I can grasp, where I am surrounded by alien and random information. This made me wonder, whether the way we look at images today—digitally, with an avalanche always at hand—prompts us to see like computers do, but without their capacity for indexation.

The formal approach I took on the university's archive made it possible to map out its enormity in a way a content-oriented approach could not. The random links formal criteria build between a mass of eclectic images creates interesting, unexpected connections. ■

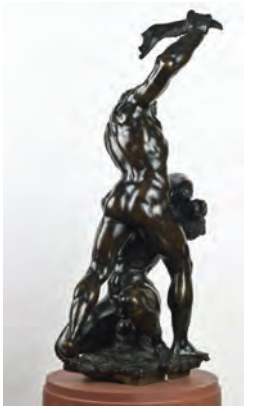
FABIENNE HESS *is an artist based in London. Documentation of the exhibition Hits and Misses and other projects can be seen on her website www.fabiennehess.com*

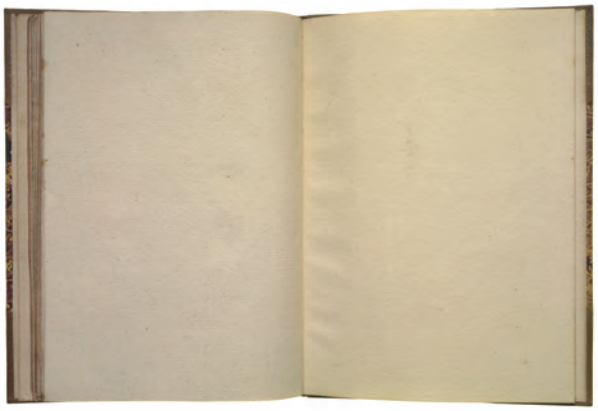
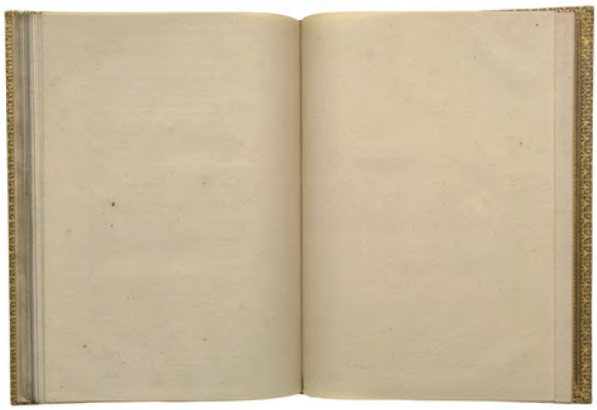
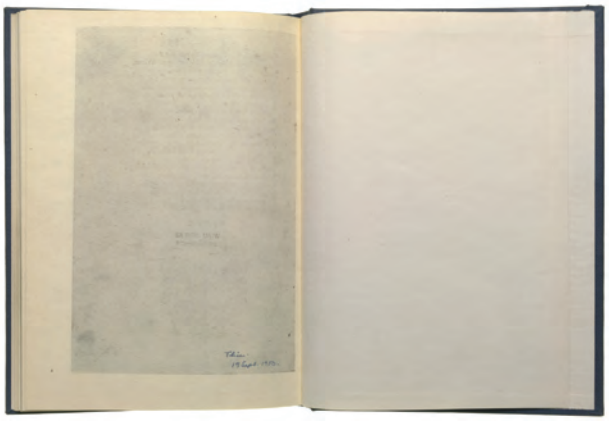
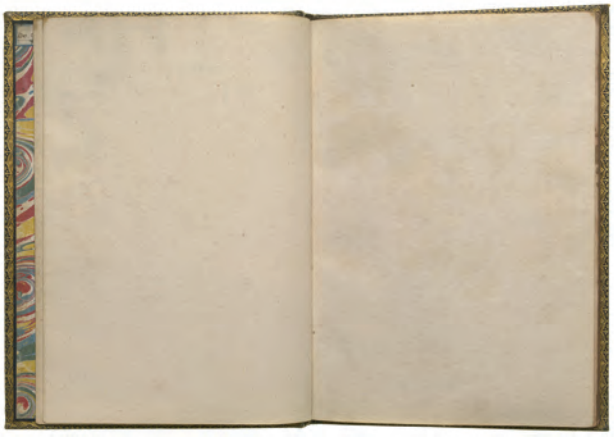
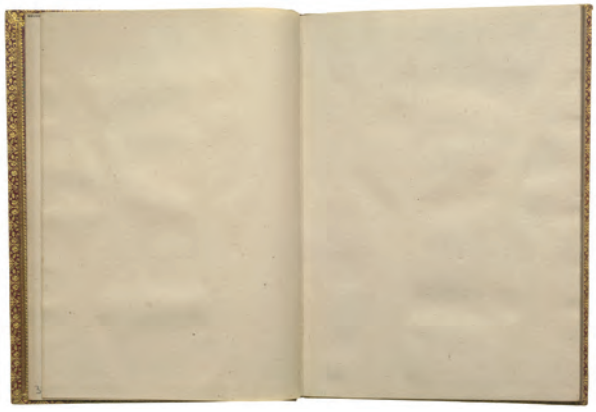
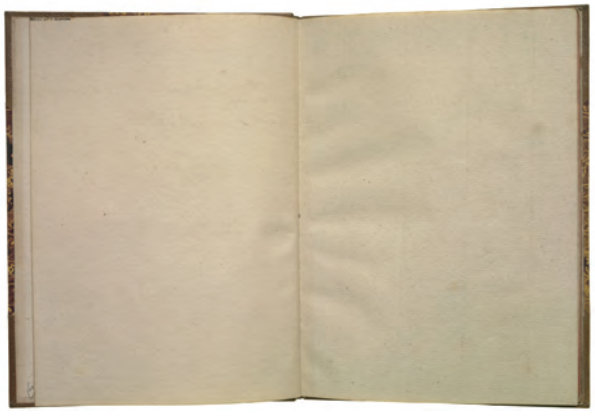
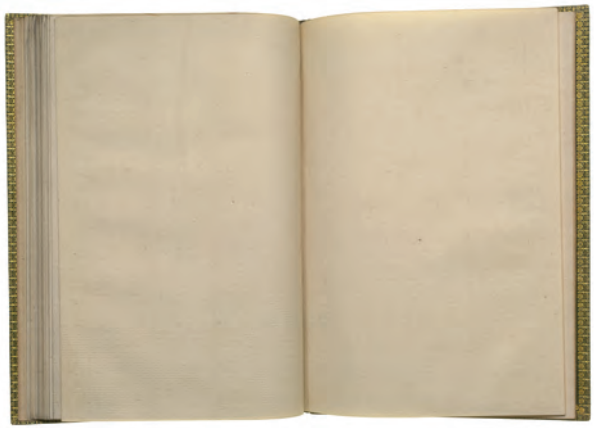
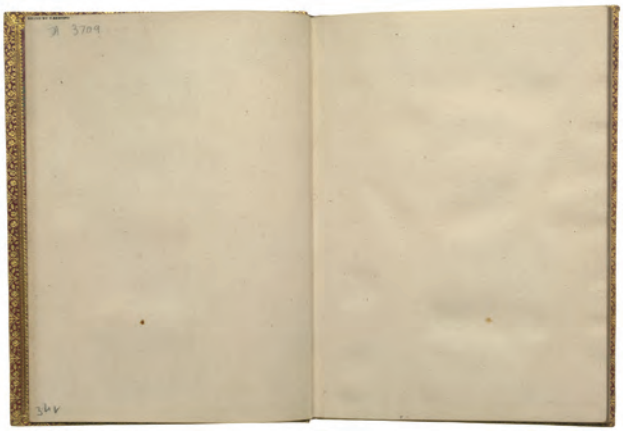

(TRIPLE POWER).

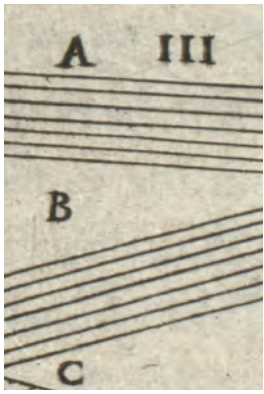


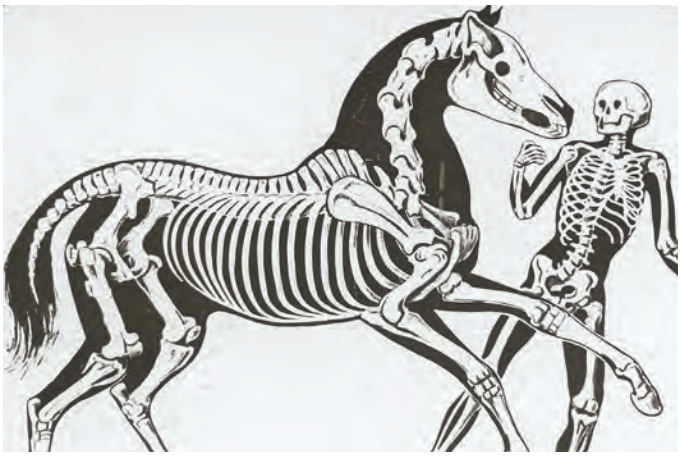

Astrachan curl


Broadtail flat moire



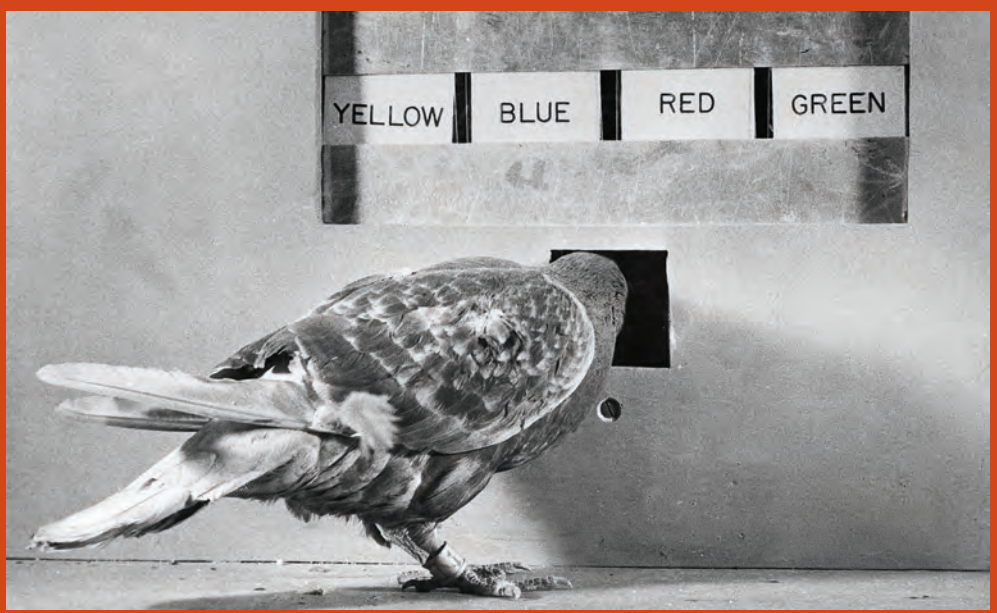








A 3709
36v


A III
B
C



just what are we archiving?

What kind of people will we become if we keep trying to archive everything? Geof Bowker reports from inside the Skinner Box.



I **WAS BORN AT THE DAWNING** of the Age of Aquarius; I have graduated to the dawn of the era of Big Data. It seems as if all aspects of our lives are being tracked, monitored, and stored for future use, and by "us" here I include vast swathes of the nonhuman as well as human world. We leave traces everywhere, often without realizing it, and these are potentially stored forever; collectively, they build a picture of ourselves that can be exploited by commercial companies (by way of Google, Facebook), governments, and aspiring political candidates.

I grew up thinking that archives were dusty, dry places that only aspiring historians such as myself could find exciting...and I still treasure the peace of roaming through a nineteenth-century set of police reports on a political group (the National Union of the Working Classes) in the 1830s in England, as well as continue to feel the anguish of what I found there. I have graduated to seeing archives as performative: they constitute the present as much as document the past.

I felt I had entered the world of the Skinner Box when I read the chilling directive by Barack Obama on September 15, 2015: "Executive Order—Using Behavioral Science Insights to Better Serve the American People" (White House 2015). For example, Section 1.b.(iii) called for scientists to: "identify programs that offer choices and carefully consider how the presentation and structure of those choices, including the order, number, and arrangement of options, can most effectively promote public welfare, as appropriate, giving particular consideration to the selection and setting of default options." At the time, I was reading Natasha Dow Schull's (2012) *Addiction by Design*, which described exactly the same logic in use by the casinos in Las Vegas to gather as much information as they could about their clients: tracking their movements, taking videos of them as they were robbed by one-armed bandits, predicting when they are about to leave so that they could be given a quick fillip to keep them going. And then as I was writing this piece, I learned that Ted Cruz had deployed Cambridge Analytica to harvest data on tens of millions of Facebook users (through the neat trick of getting Amazon Turkers to give access to their profiles, an access that extended exponentially to all their "friends" in the business) (Davies 2015). What interested me was that the government, the gambling industry, and the hopeful candidate were deploying exactly the same archival practice: find as many traces as you can of a person, track how they have responded in the past to particular messaging, then tailor your messaging so they will act appropriately in the future.

These techniques were not developed for such lofty purposes. They were part of the advertising empires built by such titans of our times as Google and Facebook. The performativity of this kind of archive is the sell: a pair of trousers, a government program, a gambling addiction, or a candidate. There is a certain beauty in the flattening of the spheres: it's all about the same archival technique. I note the resonance with James Beniger's (1986) observation in *The Control Revolution*: control is about feedback loops, which is about gleaning, siloing, and consuming information. He notes that his period, the late nineteenth century, saw the birth of mass advertising.

What is interesting about our new archival practices is that they don't want to just slot me into a category (white, elderly, middle class): they just want *me*, the traces that I have left and am leaving in the present. What is the message that works for me, Geoffrey Bowker? Let my traces tell you.... I've sure bought a lot of stuff and arguments lately I never meant to originally.

Helen Verran warned against the "hardening of the categories" (1998) what we are dealing with here is an advanced case of the softening of the categories. Lest I sound too cynical, let me say that this is an ontologically beautiful vision. A field I have found fascinating for years is that of mereology (the science of parts and wholes) because it does not assume the identity of a thing with itself over time. I share few cells and social associations with the Bowker that my mother birthed, and an ontological commitment that does not seek to put me into a finite set of fixed pigeonholes is, I think, a Good Thing. Transience, along many dimensions, is the rule: fixed categories are, to borrow a phrase from

Michel Tournier, the vitriol of the soul. Designing a fluid archival trace is a wonderful thing. The strong vision—shared by the Tardean in Latour and by Chris Anderson in *Wired Magazine*—is that our new archival practices will permit an infinitely rich view of the world that does not need to go through the annoying filter of totalizing theory.

This is a vision based on our archival traces being so complete that we can capture all that is needed to describe and act in the world. And yet the data don't even know we should be monitoring. In the world of biodiversity policy, almost all of our traces come back to the genome of a particular species. However, we are all obligate symbionts, housing within our bodies a vast array of flora and fauna and in our homes another extensive array. Yet we don't build our archives around relationships, we build them around things (if there is one fundamental flaw in our generic archival practices, it is this).

What is interesting about our new archival practices is that they don't want to just slot me into a category ... they just want me, the traces that I have left and am leaving in the present.

plenum is not yet there: it is theoretically impossible. Where we are left is inhabiting an uneasy zone between a proximate future, always just around the corner—about five years away, when we will have all the data—and a set of archival practices in the present that perpetuate certain kinds of invisibility: things we cannot or will not see.

So what kinds of things are *not* being archived in this world of total data? One kind of trace Google does not have (though it has access to all of my email, courtesy of my university) is quite simply the things I choose not to write. When I leave written traces I—along with major companies—choose to leave a record which (I hope) cannot come back to bite me. A number of years ago, a university dean advised me that an angry email I had sent was okay in sentiment, but it was not the kind of thing one said in email; one said it face to face. In the halcyon early days of the Internet, companies such as Microsoft got into trouble because they left traces of their internal strategy to scupper Apple (we saw how well that worked...). We are active contributors to the archival traces we leave, and while we may be tracked along myriad dimensions, we are not tracked along all of them. The archival traces we choose to leave provide a picture of ourselves as we would have acted had we been model citizens. A second kind of trace not being left are the traces that powerful actors work to keep out by choosing not to measure them. Thus it has taken a wonderful yearlong campaign by the *Guardian* newspaper to get the FBI to agree to put together some kind of centralized database of police killings in the United States. Things that are not being measured cannot come back to bite you: they are left scattered in unconnected databases that, in that ever-proximate future, will be linked. Despite massive data collection efforts after the Chernobyl incident, many of the effects of the disaster were not captured (Kuchinskaya, 2014), though I realize as I write this that I can only assert it, since there are no data: estimates range between a low of about five thousand and a high of mid-six hundreds of thousands. A third kind of trace is the trace that we

The dark vision of the future is that the performative archive we are creating is one destined to make Skinner right. I do indeed respond to many of the stimuli I am being prodded with, and better the data analytics on the archive, the more I respond accordingly. The bright vision (as a child of the Age of Aquarius, after all) begins with the recognition that in the world we are building—a child of the era of governmentality discussed by the early Foucault—our archives are central cultural, social, and economic constructions, and should be analyzed as rich, complex creations as much as stark, useful facts about the world. If our archives are to constitute our being, let's make them interesting. ■

GEOFFREY C. BOWKER *is Professor of Informatics at University of California, Irvine and is working on the nature of the Archive.*

BIBLIOGRAPHY

Beniger, J. R. 1986. *The Control Revolution: Technological and Economic Origins of the Information Society*. Cambridge, MA: Harvard University Press.

Davies, Harry. 2015. "Ted Cruz using firm that harvested data on millions of unwitting Facebook users." *The Guardian*, December 11. Available at http://www.theguardian.com/us-news/2015/dec/11/senator-ted-cruz-president-campaign-facebook-user-data.

Dow Schull, N. 2012. *Addiction by Design*. Princeton, NJ: Princeton University Press.

Verran, H. 1998. 'Re-Imagning Land Ownership in Australia'. *Postcolonial Studies*, Vol 1, No 2, pp 237-254.

White House. 2015. Executive Order—Using Behavioral Science Insights to Better Serve the American People. September 15. Available at https://www.whitehouse.gov/the-press-office/2015/09/15/executive-order-using-behavioral-science-insights-better-serve-american.

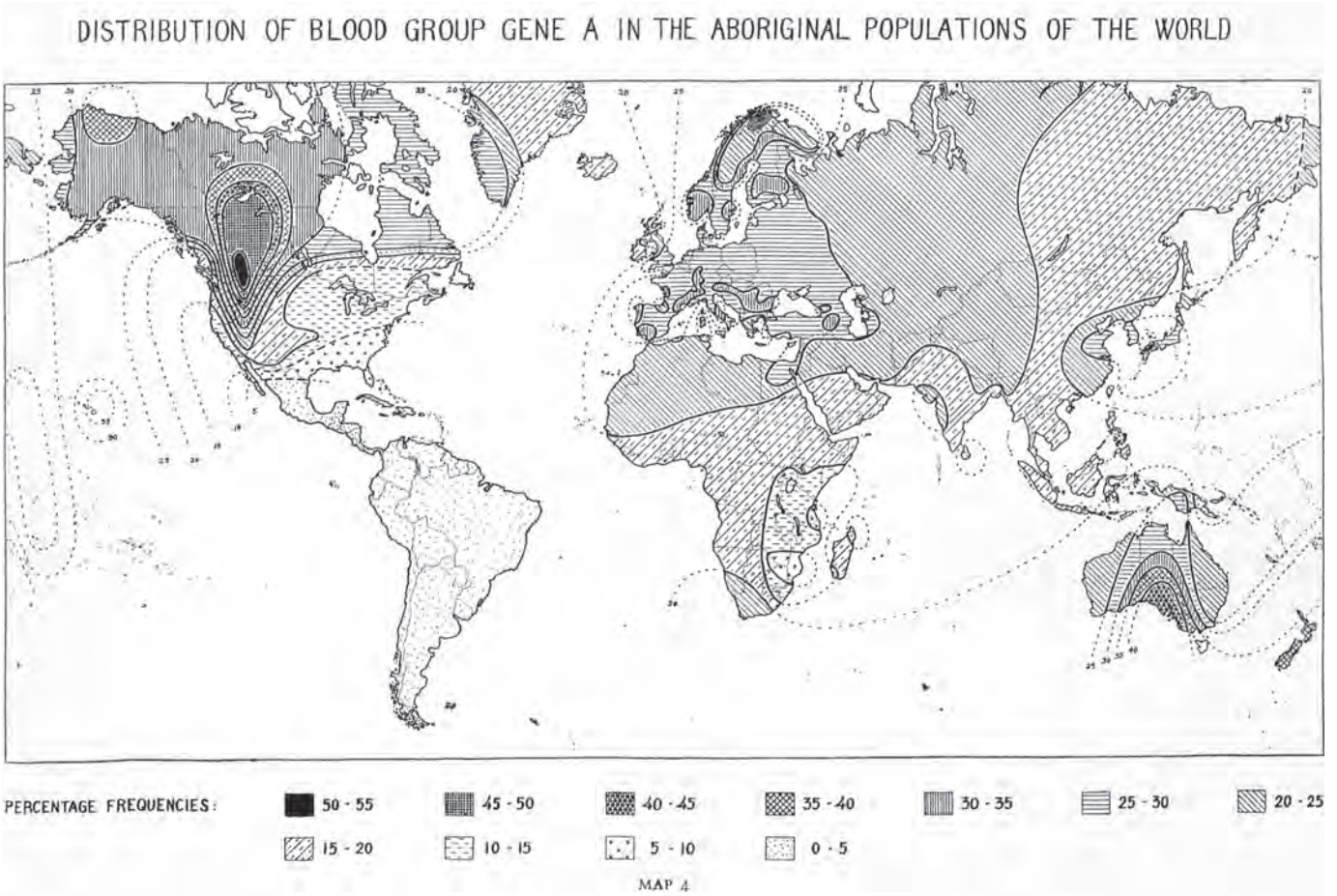
blood, paper, and total human genetic diversity

Jenny Bangham
explains how the attempt to create a supply of all possible types of human blood gave rise to genetic diversity research in the 20th century.

THE FIRST "CLEARING HOUSE" FOR HUMAN GENETIC diversity data was established in 1951 in a small building at the back of the Royal Anthropological Institute in Bedford Square, London. There, at the Nuffield Blood Group Centre, a librarian, clerk, and statistician collated and ordered a vast paper archive of blood-group data, overseen by Arthur Mourant, a hematologist affiliated with the World Health Organization (WHO). At the time, blood groups were some of the very few human traits with clear genetic inheritance, and blood-group data were being abundantly produced in the context of blood transfusion.

The humble setting of Mourant's clearing house belied its lofty ambitions. Announcing the new Centre, the U.S. magazine *Science News-Letter* claimed that blood-group data would offer nothing less than a new way of understanding human history and diversity, revealing "the genetic relationships of different groups of people" and making visible the "past nomadic wanderings and migrations of early human tribes over the face of the earth" (*Science News-Letter* 1951:237). The anthropological journal *Man* added that such data had the potential to reveal "anthropological element[s]" of disease causation and that the Centre would make an important contribution to medical research (*Man* 1951:154). To accomplish this, the workers at the Blood Group Centre would standardize and compute not only data gleaned from Britain's National Blood Transfusion Service, which had more than a million registered donors on its books, but also blood-group data extracted from published and unpublished results sent in by doctors, transfusion workers, and missionaries around the world. The Centre's first major volume—an atlas of human genetic diversity called *The Distribution of the Human Blood Groups* (1954)—represented data collected from more than 50 countries (Figure 1).

Mourant's archival ambitions were made possible by his practices as head of the Blood Group Reference Laboratory a few miles down the road in the London borough of Chelsea. Blood groups are inferred by testing blood samples against antibodies (antisera) extracted from the blood of human donors; Mourant's Reference Laboratory made and distributed standardized antisera to hospitals and transfusion centers around the world. Established in 1946 as part of Britain's peacetime blood transfusion service, Mourant's lab was designated the central blood-grouping laboratory of the WHO, which Mourant perceived to be a golden



opportunity for the large-scale collation of blood-group data. In letters that accompanied bottles of serum to colleagues in his extensive WHO network, Mourant persuaded his correspondents to send to him the results of blood-group tests. He channeled the administrative power of an organization with totalizing political ambitions (the WHO) towards the scientific rearticulation of human diversity in genetic terms.

The "wet lab" work of circulating blood samples around the world allowed Mourant to construct the "dry" archive of paper records at the Nuffield Blood Group Centre. The contribution he was making through the expert management of physical samples could be used as a bargaining chip in the collection of "anthropologically" useful data. Although novel in its scope, and in the particular nature of the deal Mourant struck, his dual project put into a formal arrangement a link between human genetic research and the procurement of blood that had been developing for thirty years. Blood transfusion had first become an indispensible surgical practice during the First World War, when techniques for preserving blood had made possible its transfer between bodies. When the war ended, increasing demand for blood forced hospitals to develop strategies for recruiting and managing donors. In some places, hospitals and private agencies paid "professional" donors high prices for blood; in others, organizations such as the Red Cross carefully fashioned donation as an altruistic service (Swanson 2014). Everywhere, transfusion produced an expanding bureaucracy to ensure a reliable supply of blood.

It was in this (still fragmented) bureaucracy that the ABO blood groups (Figure 2) accumulated as written objects of record on donor lists and cards. Their availability meant that geographically specific blood-group frequencies joined skulls and skin in hundreds of studies of racial, religious, and national differences (Schneider 1996). But blood group documentation also produced novel kinds of research. Blood groups had simple inheritance, donor records were discrete and mobile, donor lists could be collated and reproduced, and "populations" of blood groups could be summarized and repurposed. A small group of researchers intent on remodeling genetics as a population science used these properties to turn blood groups into genetic objects. To British scientists R. A. Fisher and J. B. S. Haldane, who were developing theoretical models for explaining evolutionary change through gene

FIGURE 1: One of nine fold-out maps in Arthur Mourant's *The Distribution of the Human Blood Groups* (1954), indicating the world frequency distribution of Rhesus blood-group allele C. In mapping the heterogeneous and patchy collections sent to the Nuffield Blood Group Centre, the workers there used shading and isolines to indicate a smooth diffusion of genetic variation.

FIGURE 2: Blood groups are determined by testing an unknown blood sample against specific antisera.

dynamics, blood-group data produced a vision for how these practices might be applied to human heredity. They believed that blood-group data would transform a field concerned with the inheritance of rare complex diseases into a science of large quantities of data and rigorous mathematics, thereby producing a firm basis for curtailing the propagation of genetic disease in future generations (Mazumdar 1992). This was to be—as Fisher put it in a grant application to the Rockefeller Foundation in the mid-1930s—a "solidly objective" human genetics under "strict statistical control" (Fisher 1934).

The brutality of the First World War had made people into resources for procuring blood, but the Second World War produced the conditions for the large-scale, centralized management of donors, which in turn became a plentiful resource for geneticists. The transformation from a local to nationwide transfusion service was particularly dramatic in Britain: during the first few months of the war, fragmented donor lists were brought together and remodeled within the country's first national health service: the wartime Emergency Blood Transfusion Service (EBTS). To the British government, a nationwide service for distributing disembodied blood was an essential line of defense against new technologies of aerial bombardment. And when the EBTS was established, blood donation was consciously repackaged as a contribution to the war effort (Whitfield 2013): enrollment cards proclaimed, "your blood can go on active service". It was a phenomenally successful campaign. Barely two months after recruitment began, *The Times* announced that the service had registered its first 100,000 donors (1939). This offered the potential for quantities of genetic data on a hitherto unimagined scale. Fisher's and Haldane's vision for human heredity research continued to grow. For instance, when war broke out, Fisher's lab was co-opted by the Medical Research Council and turned into the central blood-grouping laboratory for the EBTS. He took that opportunity to use the lab's contacts with depots across the country to acquire hundreds of thousands of donor records as a resource for mapping the genetic diversity of the British people, a precursor of Mourant's later, international project.

In the years following the end of the war, Britain's Ministry of Health attempted to standardize—right down to the level of typography—the management of blood and people (Figure 3). As transfusion was scaled up, more and more blood groups were discovered, and the *specificity* of blood became a new focus of bureaucratic concern. With so many people on its registry the new, peacetime National Blood Transfusion Service had reliable supplies of the common blood types, and it became increasingly focused on donors with unusual blood. While the "search for rare blood" became a dramatic narrative theme in films, plays, and newspaper reports, Arthur Mourant—by then one of the principal authorities within the transfusion service—oversaw the production of a new bureaucratic technology: a nationwide "rare blood panel" comprising a list of 2,000 donors with the rarest blood types. If a hospital anywhere in the country needed rare blood for a patient, it would telephone Mourant's laboratory in London and consult the nationwide panel for a match. Only with large numbers of registered donors in a standard nationwide service was the specificity of rare blood made visible.

The specificity of blood types became sharper as the donor registry became larger. Whereas in the 1920s a person could be A, B, AB or O, by 1950 a patient could be identified by six separate blood-group systems, of which perhaps only the Rhesus system has joined ABO in the popular understanding of blood. Mourant used this increased specificity in his anthropological archive: the greater the quantity of data he could accrue, the more detailed his geographic maps of human genetic diversity. The WHO's ambitions to connect up and standardize transfusion services around the world gave him reason to believe that the data could be collected in perpetuity. Moreover, the vastness of the collection would guarantee its own objectivity. Mourant himself admitted that he could not attest to the trustworthiness of many of his correspondents, but he judged that it was more important to encourage the free sending of data than it was to be too picky about technique, and reasoned that any errors in this heterogeneous collection would be swamped as long as he was able to collect large enough quantities of data. Even with data comprising tests on two million people, Mourant believed the collections had to continue.

UNESCO endorsed the value of blood-group-based population genetics to an international public in a high-profile campaign to undermine racial prejudices (and assert its own universalizing authority) through the dissemination of "scientific facts." Launched in the late 1940s, the race campaign was premised on the notion that a clearer scientific understanding of race among the general public would undermine prejudice. Some UNESCO experts insisted that 'race' was, in a biological

sense, 'real', but that it provided no basis for superiority or prejudice. This strand of the campaign seized upon *genetics* as a pre-eminent example of a neutral, reforming, universalizing science. Blood-group gene frequencies – the argument went – affirmed the existence of biological differences between human populations, but also flattened and neutralized racial hierarchies: for UNESCO they were the perfect mediators of racial difference. Moreover, the kind of endeavor carried out by Mourant – to map blood-group frequency diversity and thereby produce a picture of human history – was highlighted as proof of the virtues of taking a population-genetic approach to race. Dovetailing with UNESCO's commitment to "unity in diversity," the study of human population dynamics promised access to deep commonalities that tied the peoples of the world together (see Jardine, Laemmli, and Kaplan in this issue). This argument was encapsulated in the cover design of UNESCO's 1952 picture book aimed at children learning in school classrooms, *What Is Race?* (Figure 4).

In the 1950s, the wet blood/dry paper dichotomy began to break down. In the middle of that decade Danish physician Knud Eldon invented a technology that combined blood *with* paper. Blood grouping using 'Eldon Cards' involved applying blood samples directly onto a card impregnated with antibodies. With an agglutination pattern preserved on the card itself, it could be filed as a true and original permanent record. Eldon Cards were highly controversial: many doctors believed that the "apparent simplicity" of the cards would have "disastrous" consequences for patients, yet they would continue to be marketed for use in domestic settings (Figure 5). More widely enduring was the 'Guthrie' card, invented in the early 1960s by US clinical microbiologist Robert Guthrie for testing newborns for the genetic condition phenylketonuria. Today, Guthrie Cards are still routinely used to collect the blood of newborns for an array of protein and genetic tests. Despite restrictions on the length of time cards can be kept, their superlative archival qualities are being affirmed by new research programmes to repurpose the

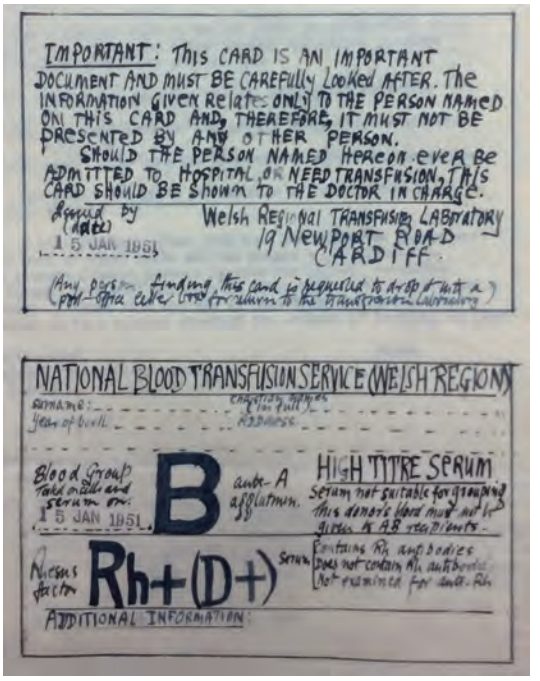

IMPORTANT: This CARD IS AN IMPORTANT DOCUMENT AND MUST BE CAREFULLY LOOKED AFTER. The INFORMATION GIVEN RELATES ONLY TO THE PERSON NAMED ON THIS CARD AND, THEREFORE, IT MUST NOT BE PRESENTED BY ANY OTHER PERSON.
SHOULD THE PERSON NAMED HEREON EVER BE ADMITTED TO HOSPITAL OR NEED TRANSFUSION, THIS CARD SHOULD BE SHOWN TO THE DOCTOR IN CHARGE.
Issued by (date) 15 JAN 1951
Welsh Regional Transfusion Laboratory 19 Newport Road CARDIFF
(Any person finding this card is requested to post it to the Transfusion Laboratory)
NATIONAL BLOOD TRANSFUSION SERVICE (WELSH REGION)
Surname:
Christian names (in full)
Address
Year of birth
Blood Group Tested on cells and serum on: 15 JAN 1951
B
anti-A agglutinin.
HIGH TITRE SERUM
Serum not suitable for grouping this donor's blood must not be given to A B recipients.
Rhesus factor
Rh+(D+)
Serum
Containing Rh antibodies
Does not contain Hr antibodies
Not examined for anti-Rh
ADDITIONAL INFORMATION:

FIGURE 3: A suggestion submitted by the Cardiff Blood Transfusion Depot to the Ministry of Health for a new standard design of blood donor cards, January, 1951. NATIONAL ARCHIVES BN 13/65.

paper-based blood spots. By preserving blood on paper the Eldon and Guthrie cards both offered an authentic record of the biochemical specificities of blood and gave it new archival possibilities.

In the 1960s several other technologies further disrupted Mourant's wet/dry economy. First, blood was refracting into an array of new protein polymorphisms: techniques such as gel-electrophoresis (separating proteins using an electrical charge) revealed hemoglobin and enzyme variants that were, like blood groups, genetically inherited. Human chromosome preparations—also made from extracted blood—gradually became a compelling new area of research. Second, novel technologies of cold storage that made possible a new material form: the *freezing* of blood samples. Whereas in the 1940s blood-grouping tests had been possible only on freshly extracted blood, now protein polymorphisms could be resolved from

FIGURE 4: Cover spread of the book *What Is Race? Evidence from Scientists* (Paris: UNESCO, 1952).

freezeable samples (Radin 2014). And not only *known* genetic variation: frozen serum was stable enough to be kept for genetic tests that might be discovered in the future. Whereas Mourant's wet/dry bureaucracy had attempted to manage an ever-increasing store of one kind of information, now blood itself, with its apparently unlimited potential, could be archived in frozen form, prompting enterprises such as the large-scale blood collection projects of the International Biological Programme (Radin 2013). Through cards, freezers and new genetic markers the anthropological quest for genetic diversity data began to diverge from the medical pursuit of stable and reliable supplies of blood.

The combination of wet and dry scientific administration that Mourant made so productive in the early 1950s captured a particular moment of alignment between wartime regimes of blood donation, postwar internationalism, and a population genetics promoted by reformists and technocrats. But several features of the postwar moment have endured. One characteristic of Mourant's enterprise that has persisted to the present day—and which has only expanded its persuasive power—is the justification of studying population differences as a way of better understanding disease. From the HapMap project to the Indian Genome Variation Consortium to the Mexican Genome Diversity Project, initiatives to collect blood for mapping human variation have promised powerful insights into disease causation. Related, large-scale genetic mapping projects project a notion of 'world citizenry' that echoes UNESCO's 1950s endorsement of genetics (notwithstanding deep present-day concerns about who benefits from such research). Finally, in spite of the severing of the institutional and infrastructural ties between transfusion and genetics, these enterprises remain linked through the social practices of donation. Civic commitments to community and nation have long been coupled to cultures of therapeutic donation, and for some these remain the framework within which blood donation for genetic research is understood (Reddy 2013). The kinds of civic responsibility that engender blood giving for genomics are still shaped by ties to community, nation and humanity that were forged through blood. ■

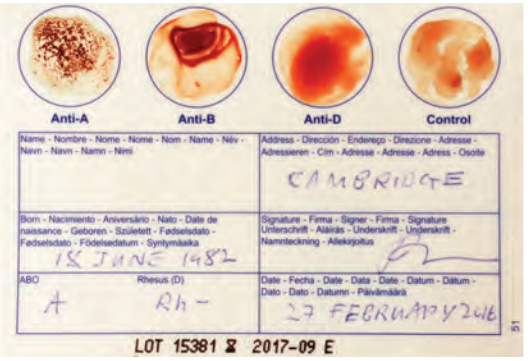


FIGURE 5: Photograph of an Eldon Card purchased on Amazon, February 2016. The test card comes with a plastic cover that can be applied once the card is dry, preserving the agglutination pattern. Pointing to the market for such tests, the same Amazon webpage listed a range of blood-group diet guides under the heading 'Customers Who Bought This Item Also Bought'.

JENNY BANGHAM *is a historian of science at the Max Planck Institute for the History of Science in Berlin, and is writing a book on mid-century blood transfusion and human genetics.*

BIBLIOGRAPHY

Fisher R.A., "To O'Brien," July 18, 1934, 01.0001/401A/Box 16, Rockefeller Archives.

Man. 1951. "Blood Group Research: Nuffield Foundation Grant to the Royal Anthropological Institute." 54:154.

Mazumdar, Pauline. 1992. *Eugenics, Human Genetics and Human Failings: The Eugenics Society, Its Sources and Its Critics in Britain*. London: Routledge.

Mourant, Arthur. 1954. *The Distribution of the Human Blood Groups*. Oxford: Blackwell Scientific Publications.

Radin, Joanna. 2013. "Latent Life: Concepts and Practices of Human Tissue Preservation in the International Biological Program." *Social Studies of Science* 43:484–508.

———. 2014. "Unfolding Epidemiological Stories: How the WHO Made Frozen Blood into a Flexible Resource for the Future." *Studies in History and Philosophy of Biological and Biomedical Sciences* 47:62–73.

Reddy, Deepa S. 2013. "Citizens in the commons: blood and genetics in the making of the civic." *Contemporary South Asia* 21:275–290.

Schneider, William H. 1996. "The History of Research on Blood Group Genetics: Initial Discovery and Diffusion." *History and Philosophy of the Life Sciences* 18:277–303.

The Science News-Letter. 1951. "Map Man's Blood Groups." *The Science News-Letter* 59(15):237.

Swanson, Kara W. 2014. *Banking on the Body: The Market in Blood, Milk and Sperm in Modern America*. Cambridge, MA: Harvard University Press.

The Times. 1939. "News in Brief." *The Times*, August 12, sec. Col. C.

Whitfield, Nicholas. 2013. "Who is my stranger? Origins of the gift in wartime London, 1939–45." *Journal of the Royal Anthropological Institute* 19:S95–S117.

THE ORIGINS OF HAPPINESS

Boris Jardine tells the story of a little ladder intended to tell us what everyone wants. Where on the ladder are you?

THROUGH THE TWENTIETH CENTURY, measures of national income and then gross domestic product (GDP) were the means by which countries came to know and assess themselves. In the twenty-first century, we are offered an alternative: gross national happiness. Massive surveys like the Gallup World Poll offer global maps of affect and desire (Figure 1), and governments seek interventions that provide cost-effective uplift in their national tally. Where the old behaviorism posited that individuals are known by their responses, the new science of happiness judges nations and continents on the basis of numerical happiness indices. But what is gained and what is lost in moving from a financial to an affective economy? Does it, as its advocates argue, signal a new kind of politics: data-fueled, evidence-based, more humane?

One way to begin to unpick gross national happiness is to look at the source of the data, and the technique used in its collection. At the heart of the Gallup World Poll

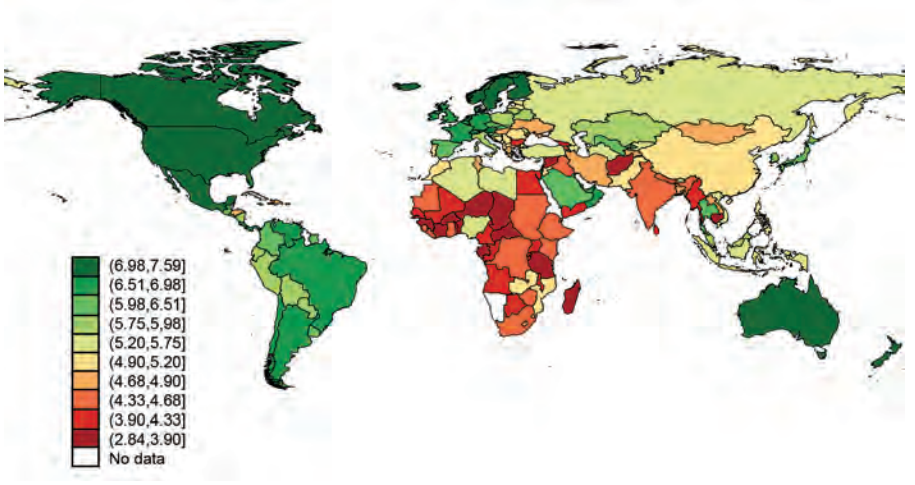


FIGURE 1: Average Cantril's Ladder scores over the period 2012–2014 (from Helliwell et al. 2015:20).

is a single question, which has been asked of people around the world for a little over half a century:

> *Please imagine a ladder with steps numbered from 0 at the bottom to 10 at the top. Suppose we say that the top of the ladder represents the best possible life for you and the bottom of the ladder represents the worst possible life for you. If the top step is 10 and the bottom step is 0, on which step of the ladder do you feel you personally stand at the present time?* (Gallup 2015).

This question is known as "Cantril's Ladder," named after the social scientist Hadley Cantril, who developed and used it on a grand scale in surveys conducted in the mid-1960s. The Ladder's most obvious feature—its simplicity—has seemed to many to be a serious flaw: its assumptions have been questioned, its phrasing refined, and other variants have been added, in which participants are asked to think forward into the future, to evaluate their lives in the round, to rank their anxiety against their contentment. Seen from another angle, however, simplicity is a virtue: no other measure of subjective wellbeing has been so widely used in a more or less stable form over such a long period of time. Cantril's Ladder provides single, numerical data points. And it has a ring of authenticity, of old wisdom. Just as Cantril intended, its assumptions are concealed beneath a deferral to the individual of whom it is asked.

But of course happiness studies do not in fact tell us anything about individual people. Nor does any single survey tell us very much either. Instead, meaning is conferred to large numbers of responses to Cantril's Ladder by studying fluctuations over time. Although happiness—under the banner of "subjective wellbeing"—is cast as an alternative to economic measures of national health, it is in fact only a refinement of economics as a practical science. The findings from happiness data that have attracted the most attention generally concern long-term correlations between traditional statistics like GDP and the newer emotional indices. Take the Easterlin Paradox, for instance, which states that although average income is positively correlated with happiness, changes in average income over time do not result in changes in happiness (Easterlin 1974). Easterlin's finding and the numerous studies that have followed rely on longitudinal happiness data from surveys using Cantril's Ladder. Up to the present day, the various data sets using the Ladder together form a temporal archive of international happiness spanning the period from the height of the Cold War to the middle of the War on Terror, via numerous economic booms and busts. This is a resource for understanding national and global trends important enough to have helped secure the 2015 Nobel Prize in economics for its most able interpreter, Angus Deaton.[1]

CANTRIL'S LADDER WAS FIRST DEPLOYED ON a global scale in the encyclopedic *Pattern of Human Concerns* (Cantril 1965). Here Cantril refined the survey techniques that he had developed as director of UNESCO's "Project on International Tensions." As Cantril put it, the Project set out to investigate

> *the distinctive features of each country's culture and ideals with view to promoting within each nation sympathy and respect for the ideals and aspirations of the others, and a just appreciation of their problems* (quoted in Rangil 2011:23).

Yet like so many mid-century projects in the social sciences, acknowledgment of cultural diversity was bound up with the search for human universals (see Banghamp. 43, Laemmlip. 59, and Kaplan p. 64 in this issue). Recording cultural difference—increasingly an end in itself for UNESCO—was for Cantril only a means with which to establish commonalities that would bind people not just to their cultural or socioeconomic grouping, but to the mass of mankind. On the one hand, Cantril complained that "cultural apolo-

1 See http://www.nobelprize.org/nobel_prizes/economic-sciences/laureates/2015/deaton-facts.html.

gists… in their search for differences between 'cultures' and uniformities within a single 'culture,' tend by and large to gloss over or explain away the individual differences which discerning readers of their own works can pick out' (Cantril 1947:21). On the other, the project of mollifying "tensions" would be superseded when social scientists moved to "a different, a 'higher,' order of accounting which must include man's desire to develop" (Cantril 1949:365). Both of these problems—of capturing the wishes of individuals while moving to a "higher order of accounting"—could be solved, thought Cantril, if information could be gathered, stored and made accessible in just the right way.

Cantril worked on two fronts to solve the riddle of the individual and the masses. The first was the establishment of a clearing house of social survey data, the Office of Public Opinion Research, whose findings were published in a bafflingly comprehensive tome in 1951 (Figure 2).

Cantril's second major project began in 1955, when he and his collaborator Lloyd Free founded the Institute for International Social Research with a huge grant administered by the Rockefeller Foundation, but in fact provided by the CIA (Simpson 1994). Over the next decade, Cantril and Free collected data from around the world using the "self-anchoring scale" that would come to be known as Cantril's Ladder. If the Office of Public Opinion Research was of interest to, as Cantril put it, "historians, sociologists, political scientists, economists, editors, policy makers, businessmen, labor leaders, and host of others," (Cantril 1951:v) then the data collected in *Pattern of Human Concerns* was more narrowly focused on the role of public opinion in statecraft. In essence, the Institute for International Social Research was a continuation of Cantril's government work, which had begun on the eve of World War II and seen him occupy various roles in the Office of War Information, U.S. intelligence operations, and as an adviser to Eisenhower and Kennedy. Indeed, the copy of *Pattern of Human Concerns* that I used in researching this essay is, appropriately enough, inscribed by Cantril to the Johnson-era Attorney General Nicholas Katzenbach (Figure 3).

In terms of survey technique, the use of the ladder question allowed Cantril to shift the burden of judgment wholesale onto the participants in his survey. Interviewees, in responding to Cantril's Ladder, would bring the entirety of their lived experience to bear on the answer; the interviewer would add none of his or her own categories to the discussion. This would allow data collection on a vast scale while paying respect to "the point of view of individual participants" (Cantril 1965:7). But just as he handed out the work of judgment to participants in the survey, Cantril hoarded back the interpretative work for himself and his academic colleagues (not to mention his government contacts).

To grasp the full force of what Cantril was proposing, consider the other kinds of inquiry into happiness that had been conducted before *Pattern of Human Concerns*. In 1938, for instance, the British social survey group Mass-Observation set out in to discover what this thing called "happiness" was. Working in the northern English town of Bolton, they enlisted the social scientist and local celebrity John Hilton as a "judge" and sent out a leaflet inviting definitions of happiness, literary talent notwithstanding (Figure 4).

The results were sometimes amusing, sometimes trite, sometimes strangely moving ("Happiness means to live content, to seek refinement rather than fashion, to be worthy, to think quietly, talk gently, act frankly, to listen to birds, to watch and study stars, to take interest in children"). But the survey was haphazard: it relied on the circulation of the leaflets and the will of the respondents. It was too discursive, inviting unquantifiable

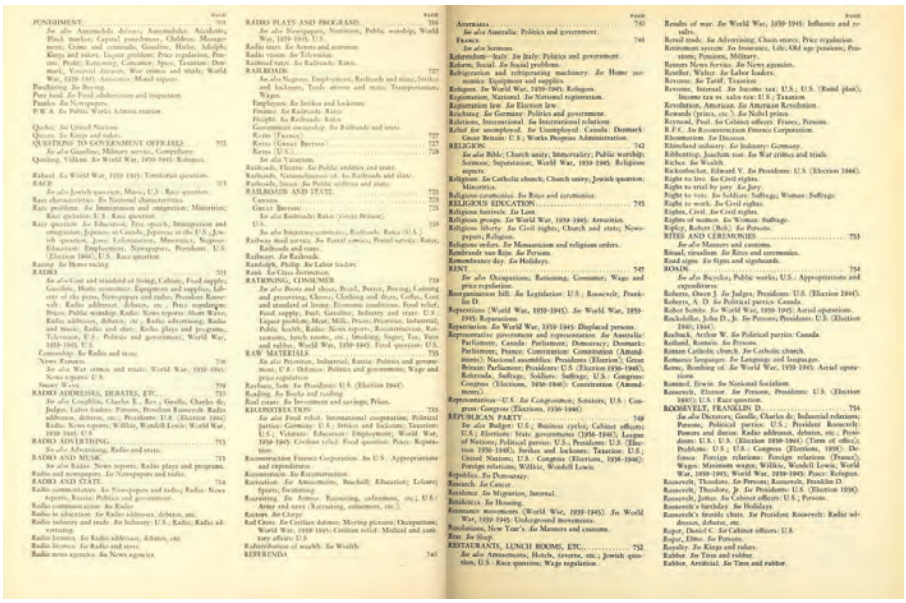


FIGURE 2. "Punishment" to "Rubber, Artificial" in the index of Public Opinion 1935–1946 (Cantril 1951).

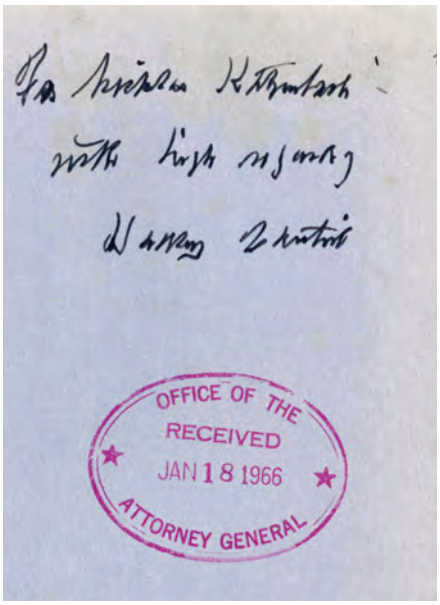


FIGURE 3. "For Nicholas Katzenbach / with kind regards / Hadley Cantril," inscribed in a copy of Pattern of Human Concerns. Katzenbach was Attorney General and then Undersecretary of State in the Lyndon Johnson administration.

data. Tabulating the results was laborious because the categories had to be extracted from digressive prose; the happiness study was not repeated.

Cantril's own book *Public Opinion 1935–1946* summarized other haphazard attempts to uncover the origins of happiness: in 1940, the American Institute for Public Research asked people to choose between "wealth, a happy home, or an interesting job" (68% wanted a happy home); in 1943, *Fortune* magazine asked young women across America whether "the next ten years of your life will be exciting ones, just average, or rather dull," giving separate scores for "single women" and "opinions of unattractive women"; in 1946, Canadians were asked whether they would rather live "in the present" (57%), "100 years from now" (13%), in the "Gay Nineties" (12%), or in a range of other eras going back to the "time of Christ" (3%) (Cantril 1951:280–281). While Mass-Observation's 1938 happiness survey shows how hard it is truly to ask people what they want, the subsequent surveys Cantril cited compare very unfavorably with the elegant technique of *Pattern of Human Concerns*.

For all its ingenuity, the deployment of Cantril's Ladder in the field was by no means straightforward. It turns out, for instance, that the idea of climbing a ladder has different connotations to different people in different places (aspiration versus labor, mundane versus exciting). Hence, in some cases Cantril's Ladder has become Cantril's Mountain (though, to mix metaphors even further, climbing a mountain is not everyone's cup of tea). Far more problematic even than this, the Anglophone concept of "happiness" itself has no obvious equivalent in many languages. A partial solution was suggested by Cantril himself, who insisted that the ladder be presented as a visual cue throughout each interview and instructed interviewers to ask the question "moving finger rapidly up and down ladder" (Figure 5). But even if this solves the problem of translation, it doesn't necessarily address the underlying conceptual mistake in assuming something as hard to define as happiness might have a common meaning throughout the world.

Beyond Cantril's own work, difficulties of standardization come into play. The build-up to the question, the kind of person employed to ask it, and the context in which it is asked have all shifted from survey to survey. Sometimes both ladder and mountain disappear completely and a simple ranking of happiness from 0 to 10 is used instead. The culturally loaded implication of striving and climbing, not to mention the idea of "complete satisfaction," are almost too obviously problematic to point out. And what of the assumptions involved in conducting survey work per se? These were known to cause problems right from the start, with one early commentator pointing out that "an educated urban man asking questions" was hardly an everyday phenomenon in many parts of the world. In some countries, interviewers were thought to be covert agents working for the government and were attacked; in Panama and Nigeria, some interviewers were arrested (Zubaida 1967:212).

For Cantril these were simply obstacles, but in reality they reflected deep suspicions that could lead to biased responses. Given that so many happiness studies are based on longitudinal data, it also suggests we should be weary of what the numbers conceal. Cantril's Ladder may appear simple, but the data have built into them the mundane, changeable, and even violent circumstances of their origin.

Over the last 50 years, the problems with Cantril's Ladder have been far outweighed by its archival fecundity. The simplicity and apparent neutrality of the technique have guaranteed its reuse, boosted by temporal studies like Easterlin's, which uncovered the income/happiness "paradox." The historicity of happiness data has led, just as Cantril anticipated, to a profound confidence in the combination of economic and social data, even as the latter purports to supplant the former in the guise of gross national happiness.

In addition to this doubling-down of economic authority, there is a more subtle consequence of using data gathered with Cantril's Ladder. The way ladder scores have been used equates individual and collective (typically national) fortunes. The more successful this identification, the weaker happiness data is in helping policymakers distinguish between alternative paths. For instance, employment tends to come up as an important factor in happiness, so a policy of full

FIGURE 4. Leaflet advertising Mass-Observation's happiness survey, 1938 (Mass-Observation Archive, University of Sussex, Topic Collection 7, Happiness 1c). Reproduced with permission of Curtis Brown Group Ltd, London on behalf of The Trustees of the Mass Observation Archive.

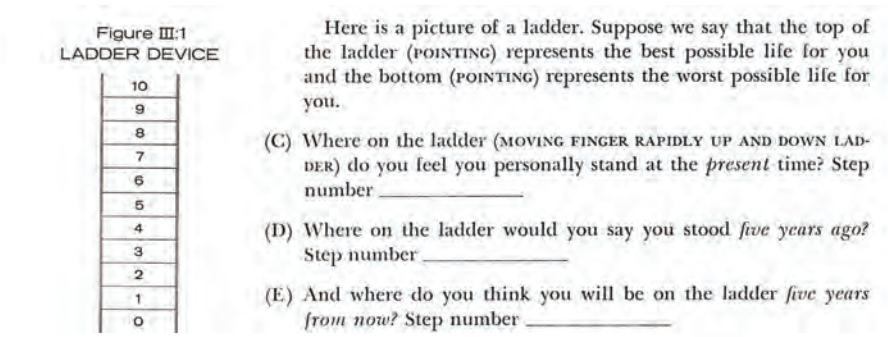


FIGURE 5. Cantril's Ladder, as illustrated in Pattern of Human Concerns (Cantril 1965:22–23).

employment seems justified, but this says nothing about wages, and so in the United Kingdom the government has pursued schemes that allow companies to circumvent the minimum wage. The cynical view is that we are moving from a politics of collective bargaining and social welfare to one of malign policy enacted on the basis of cherry-picked data purporting to represent our innermost desires (Hayward 2012). In economic terms, "balancing the books" is said to be the same for a nation as it is for a household, and so state services and local councils are dismantled with the stated aim of reducing the national deficit, the underlying aim being to boost the private sector. The "happiness industry," as William Davies has recently called it, practices the same sleight of hand in linking personal reports to national fortune (Davies 2015).

Cantril would have been fascinated by these developments; it was part of his program, after all, not only to discover human universals but to sell the idea of a highly aspirational, internationally uniform identity to the masses. If the project of *measuring* happiness were to become pervasive, then a standard of global aspiration would emerge, bringing people from diverse backgrounds together into a worldwide struggle for a single brand of well-lived life. Speculative as this sounds, there is evidence that Cantril's vision has come to pass. In a paper cited enthusiastically by the Nobel committee, Angus Deaton argues that "when asked to imagine the best and worst possible lives for themselves, points 10 and 0 on the scale, people use a *global* standard. Danes understand how bad life is in Togo [which typically scores lowest in happiness surveys]...and the Togolese, through television and newspapers, understand how good life is in Denmark or other high-income countries" (Deaton 2008:69–70).

The strange simplicity of Cantril's Ladder, the pattern of its deployment, and the interpretation of its results have brought social science, economics, and governance into close alignment. Far from displacing older economic measures, gross national happiness brings economics to bear on attitudes, on aspirations, on subjective states. The logic of totality and the power of a temporal archive have broken down the classic sociological distinction between individual and collective, and all by way of a single question. ■

BORIS JARDINE *is a researcher at the University of Cambridge, in the Department of History and Philosophy of Science.*

BIBLIOGRAPHY

Cantril, Hadley. 1947. "The Place of Personality in Social Psychology." *Journal of Psychology* 24:19–56.

———. 1949. "Toward a Scientific Morality." *Journal of Psychology* 27:363–376.

———. 1951. *Public Opinion 1935–1946*. Princeton, NJ: Princeton University Press.

———. 1965. *The Pattern of Human Concerns*. New Brunswick, NJ: Rutgers University Press.

Davies, William. 2015. *The Happiness Industry: How the Government and Big Business Sold Us Well-Being*. New York: Verso Books.

Deaton, Angus. 2008. "Income, Health, and Well-Being around the World: Evidence from the Gallup World Poll." *The Journal of Economic Perspectives* 22(2):53–72.

Easterlin, Richard. 1974. "Does Economic Growth Improve the Human Lot? Some Empirical Evidence." In *Nations and Households in Economic Growth: Essays in Honor of Moses Abramovitz*, edited by Paul A. David and Melvin W. Reder (pp. 89–125). New York: Academic Press, Inc.

Gallup. 2015. "Understanding how Gallup Uses the Cantril Scale." Available at http://www.gallup.com/poll/122453/understanding-gallup-uses-cantril-scale.aspx.

Hayward, Rhodri. 2012. "The Politics of Happiness." *The History of Emotions Blog*, Queen Mary University of London, Centre for the History of Emotions. https://emotionsblog.history.qmul.ac.uk/2012/04/the-politics-of-happiness

Helliwell, John, Layard, Richard, and Sachs, Jeffrey, editors. 2015. *The World Happiness Report 2015*. http://worldhappiness.report/wp-content/uploads/sites/2/2015/04/WHR15-Apr29-update.pdf

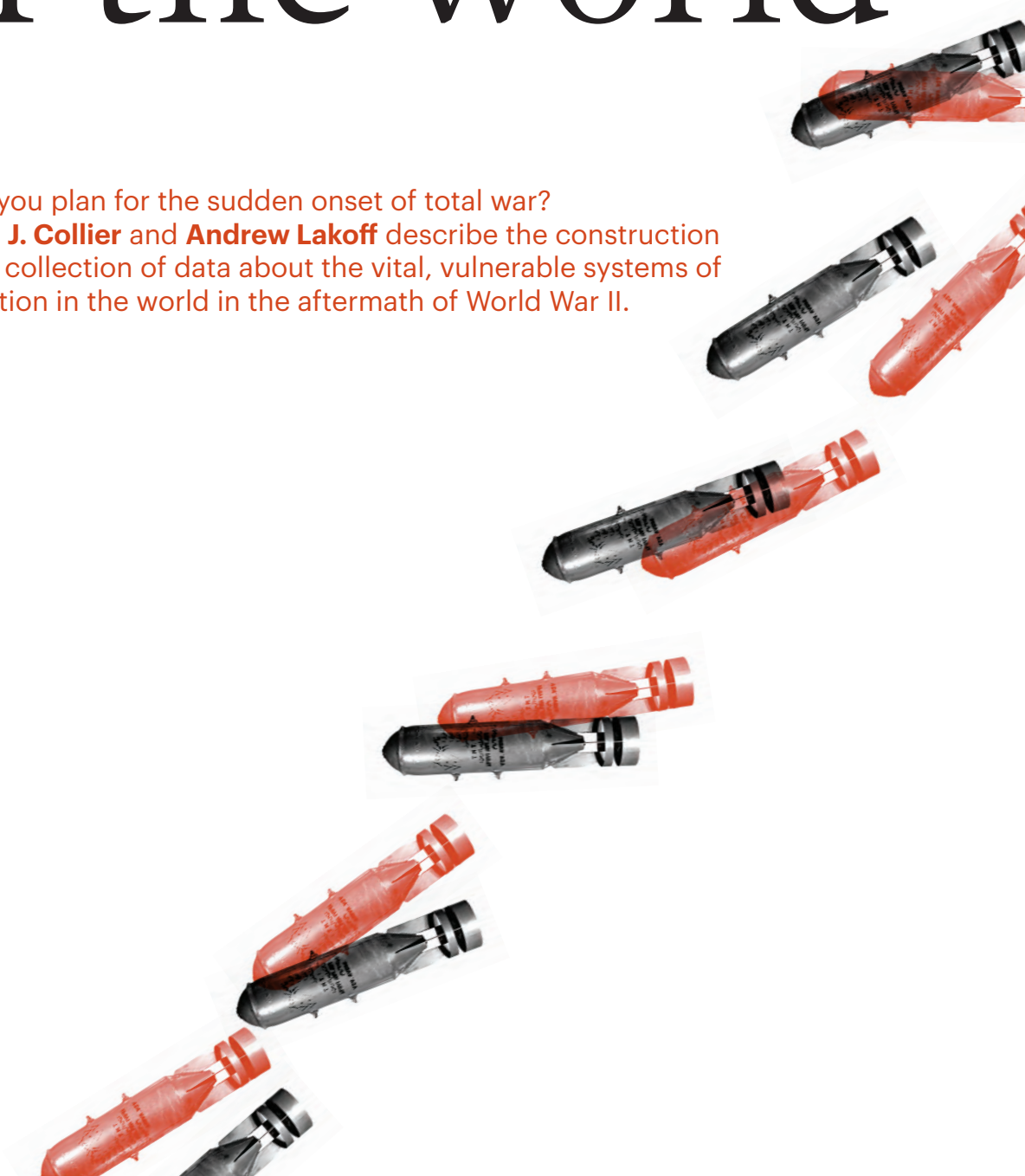
Rangil, Teresa. 2011. *The Politics of Neutrality: UNESCO's Social Science Department, 1946–1956*. Durham, NC: Duke University Center for Political Economy.

Simpson, Christopher. 1994. *The Science of Coercion: Communication Research and Psychological Warfare 1945–1960*. Oxford, UK: Oxford University Press.

Zubaida, S. D. 1967. "Review of *The Pattern of Human Concerns*." *The British Journal of Sociology* 18:212.

The bombing encyclopedia of the world

How do you plan for the sudden onset of total war? **Stephen J. Collier** and **Andrew Lakoff** describe the construction of a vast collection of data about the vital, vulnerable systems of every nation in the world in the aftermath of World War II.



IN A SERIES OF LECTURES delivered to the Air War College in the years after World War II, Dr. James T. Lowe, the Director of Research for the Strategic Vulnerability Branch of the U.S. Army's Air Intelligence Division, described the elements of "The Bombing Encyclopedia of the World." The Bombing Encyclopedia, he said, was "a huge index of basic, factual information that is being built up on all potential objectives of air attack throughout the world" (Lowe 1948:6). Lowe explained how it was being constructed. First, potential bombing targets were plotted on map inserts, and each target was assigned a number. Then, information on each target—"all necessary detailed facts that will be needed later to make a target analysis"—was punched into an IBM card using a coding system. "The combination of the maps and the codes," Lowe summarized, "is what constitutes the Bombing Encyclopedia" (1946:13). Using it, an analyst could make a punch card "run" on a tabulating machine (see Figure 1) that would identify potential targets based on a range of selected criteria.

The Bombing Encyclopedia was a remarkable effort to organize a massive amount of information that could be rapidly queried to identify air targets in the event of a future war. What is perhaps surprising is the nature of the bombing targets it identified, and thus of the data being assembled. These targets were not just military installations, but an array of industrial plants, service facilities, and other vital structures of the industrial economies of the world. And the information collected in the Encyclopedia did not only relate to the immediate military relevance of a particular facility, but also to the complex web of economic interconnections into which it was embedded: the number of employees, the volume of its total production, the raw material inputs it used, and the finished products it produced (see Figure 2).

The Bombing Encyclopedia's emphasis on industrial targets followed from the type of war envisioned at the time by military strategists, a "total war" fought not only by military forces, but by entire military-industrial economies and populations. As Lowe declared, "War in its final analysis is nothing more nor less than a contest of national energies" (1948:15). In a memo on air war planning written early in World War II, industrial economist Emile Despres argued that the contest must be understood as "a deadly race to bring American resources up to capacity while simultaneously reducing those in Germany" (Katz 1989:104). In the same fashion, a future war would be won by simultaneously mobilizing one's own military-industrial production system and "de-mobilizing" that of the enemy. Preparedness for an air assault to achieve this aim in a future war was the defining mission of the Strategic Vulnerability Branch (SVB). As Lowe put it, SVB sought to "determine the strategic vulnerability to attack of all the countries of the world including the United States and to recommend targets for destruction in the event of war" (1946:2). The task for strategic vulnerability analysts was to find the key target system—whether electricity, petroleum, transportation, airframe production, or some other system—"the destruction of which would cut across the totality, the entirety of the enemy's ability to defend himself" (1946:7).

The methodology for strategic target selection employed in the SVB had been developed during World War II by economists serving in U.S. air intelligence units such as the Enemy Objectives Unit, which identified strategic targets in Germany, and the Joint Target Group, which focused on Japan. In deciding which systems to target, these economists analyzed the relationship of a given target system (such as oil or transportation) to the production of strategically important end products by examining the flow of inputs and outputs through an adversary's industrial production system. They could then assess this target system's vulnerability by looking at factors such as "cushion" (available reserves), "depth" (how long it would take for the destruction of a given target to affect military supply lines), and "substitutability" (whether a targeted item could readily be replaced by another product).

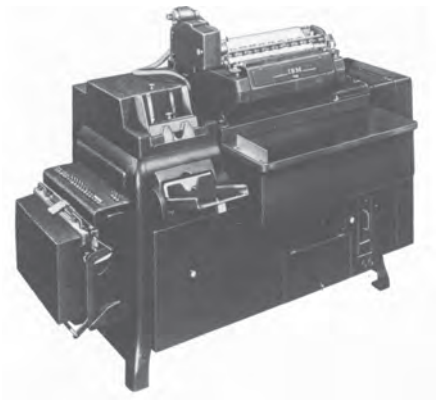


FIGURE 1. The IBM 405 Electric Punched Card Accounting Machine, a model in operation in the years immediately following World War II. In a typical application, the machine read a deck of punched cards, each of which contained a series of numeric fields. The machine then tabulated the totals for each field selected by an operator. (HTTP://WWW.COLUMBIA.EDU/CU/COMPUTINGHISTORY/405.HTML)

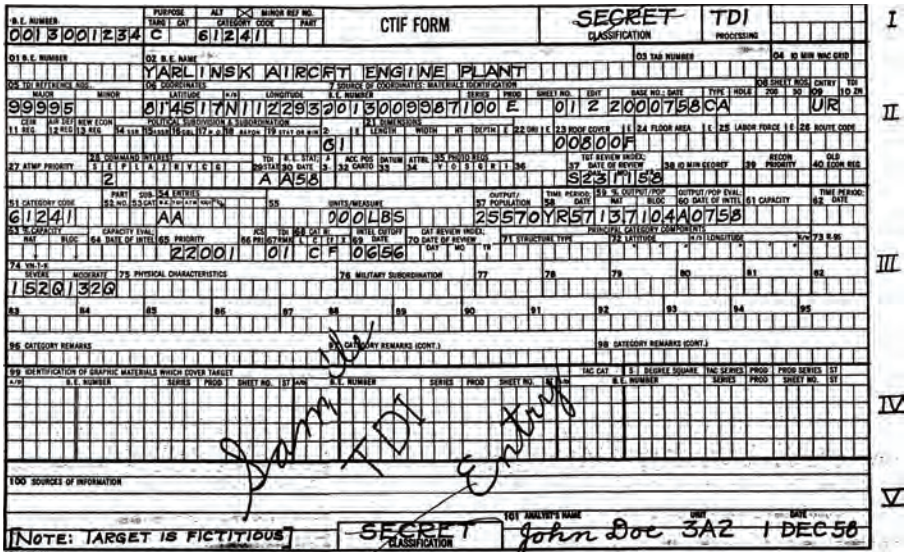


FIGURE 2. Sample of the "Consolidated Target Intelligence Form," created to help manage the information contained in the Bombing Encyclopedia. For each bombing target, a form was filled out that included: (1) codes for machine processing; (2) information identifying and locating the target; (3) information on the category of the target and its individual characteristics; (4) references to graphic coverage on the target; and (5) sources. (CLINARD 1959:100).

On the basis of these analyses, air intelligence specialists developed bombing plans for particular geographical areas that focused on the most vital facilities in a particular target system. This technique for understanding the vital flows of materials in a military-industrial economy was known as "strategic vulnerability analysis."

During the war, Lowe served as an air intelligence specialist in the Joint Target Group (JTG), where many of the techniques used in the Bombing Encyclopedia were first developed. JTG specialists compiled information on potential targets in the Japanese war economy, including the type of installation, its location, and its significance in a given target system. JTG target reports then compared potential target systems and made recommendations on optimal attack strategies. The combination of maps and codes contained in the JTG target reports (see Figures 3 and 4) was the precursor to the information contained in the Bombing Encyclopedia.

After the war, military planners asked what lessons the air war in Europe and Japan held for U.S. air strategy in the future. They pointed in particular to the lack of air intelligence early in the war, which made it difficult to determine productive air targets. "If a comparable lack of intelligence should exist at the outset of a future national emergency," argued the *US Strategic Bombing Survey*, "it might prove disastrous" (USSBS 1946:108). As Lowe put it, a future war would "begin with an attempted Pearl Harbor on the industrial heart land of the United States" (1946:12). Strategic planners sought to ensure that American air forces would be prepared in advance for such an eventuality. In this context, the SVB's mission was "to make a pre-analysis" of the vulnerability of potential adversaries around the world, and to "carry that analysis to the point where the right bombs could be put on the right targets concomitant with the decision to wage war without any intervening time period whatsoever" (Lowe 1948:4). Thus, peacetime preparedness entailed maintaining a continually updated inventory of information on the world's military industrial economies.

Recalling the experience of the JTG during World War II, Lowe noted that information about the Japanese industrial economy had not been hidden or secret. Rather, it was "open intelligence—facts about industry, location of plants, machinery that is in them, the materials that

FIGURE 3. In the latter stages of World War II, the Joint Target Group produced maps and target lists based on their analyses of critical vulnerabilities in the Japanese military-industrial economy. This target map focuses on Japanese oil production as a potential target system. It is notable that some of the facilities identified on the map are located in Japanese-occupied parts of China. Though this target system was vital to the Japanese military-industrial production apparatus, it extended beyond Japanese territory. (JOINT TARGET GROUP 1944)

10. Target list:

TARGET NO.:	NAME OF FIRM	LOCATION	IMPORTANCE (BBLS.)
90.32-673	Naval Oil Fueling Sta.	Tokuyama	Refining (2,600,000) Storage (3,282,000) Synthetic (150,000)
90.17-128	Petroleum Center	Kawasaki	Refining (2,300,000) Storage (1,000,000)
90.32-1884	Nippon Naval Oil Storage	Tokuyama-O Shima	Storage (4,320,000)
90.32-672	Nippon Oil Company	Kudamatsu	Refining (2,475,000) Storage (750,000)
90.25-1203	Nippon Oil Ref. & T.F.	Amagasaki	Refining (1,825,000) Storage (1,000,000)
90.32-669	Hanano Waki Oil Storage	Tokuyama	Storage (588,000)
90.9-90.10-1012	Nippon Oil Company	Niigata	Storage (560,000)
90.9-90.10-1013	Nippon Oil Company	Niigata	Storage (420,000)
90.17-89	Eastern Petroleum Ref.	Yokohama	Refining (180,000) Storage (300,000)
90.34-40	Asahi Oil Refinery	Hikoshima	Refining (180,000) Storage (150,000)
90.20-1684	Tokkaichi Refinery*	Utsube River	Refining & Synthetic oil (5 - 10,000,000) Storage (2,285,000)

*Requiring special bomb loading

For locations see Map JTG/1A-4.

SECRET

- 3 -

FIGURE 4. Target list from Joint Target Group Japan Report (December 1944). The target numbers correspond to some of those on the target map in Figure 3. (JTG/1A-4)

go into the plants, who uses the output, etc." (1946:12). But this information was "scattered everywhere—in military and government files, and in the files of banks, insurance companies, engineering offices, and religious organizations." Similarly, in the early Cold War the problem for intelligence analysts was not how to access hidden data, but rather "how to cope with a perfect avalanche of information." The solution, according to Lowe, depended both on having people "with sufficient professional 'know how' to handle the data when it has been attained," and on the "integrity of the plan of operations to handle this great mass of data" (1946:12-13). The analysts in the SVB would provide the relevant "know how." And SVB's plan for handling the 'mass of data' necessary for target selection was to construct the Bombing Encyclopedia.

Lowe argued that the punch card tabulation system used for the Bombing Encyclopedia would make it possible to organize and access this massive amount of data with greater efficiency, accuracy, and flexibility. The new "machine methods" of information management made it possible "to operate with a small fraction of the number of people in the target business that would normally be required." The "disadvantages of human errors and human filing systems" could be almost entirely eliminated. And the data could be flexibly accessed: it would not be organized through a single, rigid system of classification, but could be queried through "runs" that would generate reports about potential target systems based on selected criteria such as industry and location. As Lowe explained, "[b]y punching these cards you can get a run of all fighter aircraft plants" near New York or Moscow. "Or you can punch the cards again and get a list of all the plants within a geographical area.... Pretty much any combination of industrial target information that is required can be obtained—and can be obtained without error" (Lowe 1946:13-14).

The speed and flexibility of the Bombing Encyclopedia, according to Lowe, made it possible to address a new problem: the need to prepare in advance for a range of possible future contingencies, rather than for a single predetermined objective (as had been the case in World War II). Using the Bombing Encyclopedia, he explained, one could screen the 70,000 potential objectives in a given country offered through a "fine mesh" to get to "the 70 odd or 7 that are within the capabilities of the attacking air force" and whose destruction "would make the maximum contribution to the attainment of the mission" (1946:7). To illustrate, Lowe noted that a punch card run would enable the analyst to "choose between the electric power plants in a given country and the aluminum plants, the transportation system as against the steel industry, or the aircraft industry as against small arms and ammunition" (1948:13). Once such objectives had been selected, SVB could recommend specific target systems to the Strategic Air Command, which would in turn "prepare the operational target charts [to be] ready for any national emergency" (Lowe 1946:3).

DURING THE 1950S, VULNERABILITY analysts continued to add information on new targets to the Bombing Encyclopedia; by 1959 it included more than 78,000 target listings (Clinard 1959:89). Meanwhile, increasingly sophisticated technologies for data handling were developed: magnetic tapes replaced punch cards; digital computers replaced tabulators. Although the name changed (today it is known as the Basic Encyclopedia) and support for the program ebbed and flowed at different moments, the U.S. Air Force has continued to maintain something very much like the Bombing Encyclopedia up to the present day (Gregory 2012). But the broader significance of the Bombing Encyclopedia should be sought elsewhere. In developing it, strategic vulnerability analysts formalized a distinctive and novel set of techniques for organizing and analyzing large amounts of data about economic and social life, and for relating the present to an uncertain future. These techniques would have far-ranging uses beyond the specific domain of strategic target selection.

In *The Taming of Chance*, Ian Hacking (1990) showed how, in the first decades of the nineteenth century, an "avalanche of numbers" enabled a new understanding of collective existence. By recording and assembling increasingly comprehensive information about phenomena such as crime, illness, and poverty, it became possible to project the regular occurrence and distribution of past events into the future, based on the assumption that these statistical regularities would remain constant over time. A new object of knowledge and a new target of government intervention were thereby constituted: a society governed by statistical laws (Hacking 1990:46). Much like the administrative science of statistics, the Bombing Encyclopedia was built to manage what Lowe called a "perfect avalanche of information," but it was meant to contribute to a different kind of knowledge about collective existence, and to establish a very different relationship between the present and the future. The inventory assembled for the Encyclopedia was not a record of the past; rather, it was a catalog of the elements comprising a modern military-industrial economy. The analysis of strategic vulnerability did not calculate the regular occurrence of events and project the series of past events into the future, based on the assumption that the future would resemble the past. Rather, it examined interdependencies among these elements to generate a picture of vital material flows and it anticipated critical economic vulnerabilities by modeling the effects of a range of possible future contingencies. It generated a new kind of knowledge about collective existence as a collection of vital systems vulnerable to catastrophic disruption (Collier and Lakoff 2015).

A key moment in the translation of this type of knowledge from air targeting to other domains was the adoption of the techniques used in the Bombing Encyclopedia for use in the field of nonmilitary defense planning. This field was concerned not with planning attacks on enemy production systems, but with identifying and reducing sites of *domestic* vulnerability. In the 1950s, several vulnerability analysts who had served in the SVB were transferred to the Office of Defense Mobilization, the Executive Branch office responsible for non-military defense. There they were charged with developing damage assessment techniques that could anticipate the effects of an enemy attack on the American military-industrial economy. Using the new Univac computer (see Figure 5), and drawing on enormous amounts of information about industrial production, government and military facilities, and critical infrastructures, they designed increasingly sophisticated models of the effects of a projected nuclear attack on the United States. Given the concerns of nonmilitary defense—which included the survival of the civilian population in the aftermath of an attack—the range of critical systems considered in such models was expanded beyond military-industrial production systems to encompass health

FIGURE 5. UNIVAC, one of the computers used in the analysis of strategic vulnerability.

infrastructures, food systems, and public utilities that met civilian needs.

Over the following decades, these tools for modeling uncertain future events—the inventory of elements at risk, interdependency analysis to determine vulnerability, and computer-based simulations—were applied to a range of other kinds of potential catastrophes, from natural disasters, to pandemic outbreaks, to financial crises. By the beginning of the twenty-first century, such tools were understood as the key components of a generic technique for anticipating future events: catastrophe modeling (Collier 2008).

Underlying catastrophe modeling as it is applied today in many domains are practices for assembling data about economic and social life that are very similar to those used in the Bombing Encyclopedia. Systemic risks in finance are made visible through inventories of financial institutions, with their interlocking liabilities. Portfolio risks in insurance are grasped through inventories of insured properties, models of future disasters, and assessments of insurance companies' exposures. Vulnerabilities to pandemic outbreaks are understood through inventories of medical stockpiles, emergency care facilities, and medical personnel, whose interdependency is put to the test through complex models. Thus, techniques for accumulating and assessing large amounts of data initially developed in wartime air targeting continue to carve out new objects of knowledge and of governmental concern. ■

STEPHEN J. COLLIER *is Associate Professor of International Affairs at The New School.* **ANDREW LAKOFF** *is Associate Professor of Sociology and Communication at the University of Southern California.*

BIBLIOGRAPHY

Clinard, Outten J. 1959. "Developments in Air Targeting: Data Handling Techniques." *Studies in Intelligence* 3:95–104.

Collier, Stephen J. 2008. "Enacting Catastrophe: Preparedness, Insurance, Budgetary Rationalization." *Economy and Society* 37(2):224–250.

Collier, Stephen J., and Andrew Lakoff. 2015. "Vital Systems Security: Reflexive Biopolitics and the Government of Emergency." *Theory, Culture, and Society* 32(2):19–51.

Eden, Lynn. 2004. *Whole World on Fire: Organizations, Knowledge, and Nuclear Weapons Devastation.* Ithaca, NY: Cornell University Press.

Gregory, Derek. 2012. "Bombing Encyclopedia of the World." *Geographical Imaginations*, August 3.

Hacking, Ian. 1990. *The Taming of Chance*. Cambridge, UK: Cambridge University Press.

Joint Target Group. 1944. JTG Estimate No. 1, Strategic Air Employment Suitable to the Current Strategy of the Japanese War, December 1944. NARA RG18.UD21. http://www.japanairraids.org/?page_id=1932

Katz, Barry M. 1989. *Foreign Intelligence: Research and Analysis in the Office of Strategic Services*. Cambridge, MA: Harvard University Press.

Lowe, James T. 1946. "Intelligence in the Selection of Strategic Target Systems." Lecture, Air War College, Maxwell Air Force Base, Alabama. Documents M-U 38043 L913i. Muir S. Fairchild Research Information Center, Air University.

———. 1948. "The Theory of Strategic Vulnerability." Lecture, Air War College, Maxwell Air Force Base, Alabama. File K239.716250-43, Air Force Historical Research Agency, U.S. Air Force.

"Sevillana dancers in town square," FROM THE ALAN LOMAX COLLECTION AT THE AMERICAN FOLKLIFE CENTER, LIBRARY OF CONGRESS. USED COURTESY OF THE ASSOCIATION FOR CULTURAL EQUITY.

Alan Lomax and the Temple of Movement

Alan Lomax wanted to catalogue all human movement. Whitney Laemmli explores the high modern utopianism of the Choreometrics project.

IT WAS 1965: televisions were blaring, and Alan Lomax was worried. As *Gilligan's Island, The Andy Griffith Show, Green Acres*, and *Bewitched* capered into tangerine family rooms, the famous American folklorist expressed his horror.[1] TV, he proclaimed, once "promised to be a marvelous telescope that could bring the whole world into our rooms—a periscope through which we could peer, unseen and unabashed, into other lives." Instead, he continued, "it has erected an electronic curtain, composed of our own prejudices and preconceptions, through which the outside world can only be dimly perceived."[2]

This curtain was an intellectual one—it featured news coverage, for example, that neglected the concerns of the distant and the poor—but it was also something more. Lomax had begun to notice eerie changes in the way people across the planet moved their bodies, observing in particular the seeming omnipresence of "the head-back, chest-out, erect posture of the North European elite." Scholars in the new field of kinesics had found that human beings "respond below the level of awareness to the movement patterns they encounter,"[3] and he warned that the global media was in the process of invading humanity's very bones and sinews. Soon, even the denizens of the world's most remote forests would be striding like stiff-necked London bankers.

Lomax predicted that the consequences of this transformation would be profound and devastating, ranging from individual emotional pain to the wholesale destruction of embodied cultural knowledge. His solution? A massive archive of movement. Along with dance experts Irmgard Bartenieff and Forestine Paulay, Lomax set out on a mission: to collect, view, code, and catalog filmed samples of dance from every cultural group on the planet. He described his dream of:

> *...a great library of the visual arts, where all important cinematic documents would be stored, catalogued, and analyzed. Such a temple of knowledge would cost no more than an atomic submarine, but its influence would far outrun the famed library of Alexandria or, indeed, all the libraries that ever existed, since it would preserve a living, moving record of all human behavior.*[4]

This store, he hoped, would provide the raw material necessary to achieve two related objectives. First, it would provide scientific proof that cultural practices were not mere window-dressing on the human experience, but rather were crucial to human survival. While other animals depended on genetic change to produce new adaptive behaviors, Lomax believed that human beings were "revising and reorganizing" their behavior each and every day, and then passing on their collective wisdom "as parts of symbolic cultural codes, rather than as encoded in the helix of the genes."[5] He argued, for example, that the stooped posture and "deep shoulder rotation" characteristic of West African dance likely helped to communicate information about "a principal subsistence act": the use

1 For more about Lomax's biography and career, see Szwed (2011).

2 Alan Lomax, unpublished manuscript for *Dance and Human Culture*, Box 4/18-01, The Alan Lomax Collection at the American Folklife Center, The Library of Congress.

3 Lomax, *Dance and Human Culture*.

4 *ibid.*

5 *ibid.*

of the short-handled grubbing hoe. Eskimo dance, on the other hand, was characterized by rapid changes in intensity, inculcating the principle that "one effective way to generate heat in the extreme cold is to stiffen or clench a part of the body and then strongly apply energy."[6]

To provide statistical evidence for these correlations, Lomax, Bartenieff, and Paulay spent more than a decade "ransacking the film libraries of the world,"[7] reaching out to absolutely anyone who might have filmed dance. In the dozens of boxes of letters that now fill the Library of Congress archives, one can find correspondence with anthropologist Margaret Mead and kuru researcher Carleton Gadjusek, but also a retired vascular surgeon, a Russian cultural minister, the Walt Disney Company, and the U.S. military. Ultimately, the Choreometrics team gathered more than 250,000 feet of raw footage, representing 2,000 different communities. Each piece was then logged, coded for dozens of different movement qualities, and entered into a database for further analysis.

This massive undertaking was undeniably ambitious. Lomax's second objective, however, was even more so: communicating these findings—and the methodology behind them—to a global audience. In essence, he wanted to teach the world to see dance in the same way that he now did. The process would not be easy, particularly as the qualities Choreometrics coded for were not obvious to the untutored eye, but Lomax remained hopeful. In part, this optimism stemmed from his experience coaching the graduate students who helped analyze the initial collection of dances. The instructions for raters at the beginning of the Choreometric Coding Book included the following caution: "The rater is advised not to attempt to count the frequency of a feature by breaking down the action or scene into similar parts or units and then summing up his impressions in numerical terms."[8] This flight from numbers seems, at first, at odds with a project that generally

6 *ibid.*

7 *ibid.*

8 Forrestine Paulay, Irmgard Bartenieff, and Alan Lomax, "The Choreometric Coding System," Box 4/18-03, The Alan Lomax Collection at the American Folklife Center, The Library of Congress.

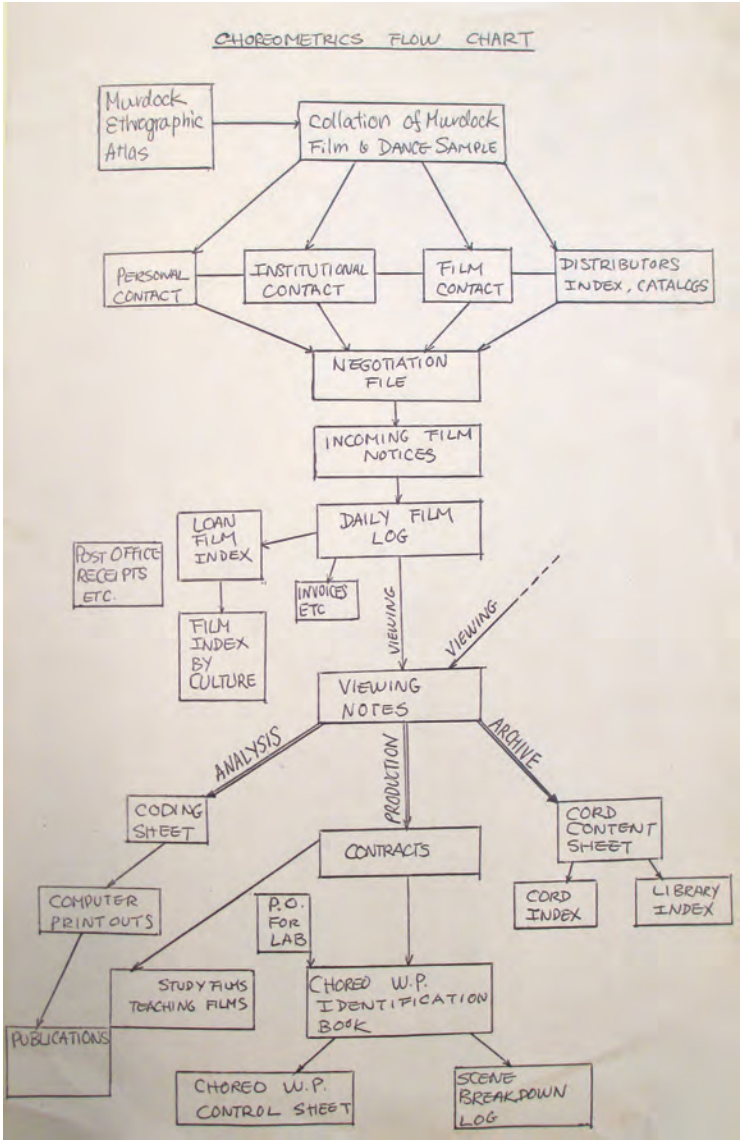
emphasized its quantitative, scientific credentials.

An integral part of Lomax's methodological theory, however, was that the system itself would ultimately render numerical measures unnecessary. As a rater viewed more and more films, coding sheets in hand, he or she would gradually absorb the schema until it became second nature. "Training in Choreometrics," Lomax contended, "consists, fundamentally, of the recalibration of the observer's standards of tempo, etc., to the full human range."[9]

Lomax hoped that such training could be extended to the audience at large, and his plans for publication reflected this radical sensibility. Though he did use the Choreometrics data to venture scholarly arguments about the relationships between dance patterns, work, culture, and physical environment, he never planned to produce an academic monograph. Instead, he envisioned a strange chimera of coffee-table book and scholarly tome. Not only would the book feature significant excerpts from the raw data, it would include coding instructions and blank versions of the coding sheets: an invitation to a DIY education in movement observation. In addition, Lomax hoped to make the films he drew upon widely available, whether broadcast on American public television—as he did with *Dance And Human History* (1975) and *Palm Play* (1977)—or screened privately to the far-flung communities they depicted. More than earlier salvage projects—which aimed to capture a past that would inevitably vanish—Choreometrics was future-oriented, dedicated to the active propagation of diversity and open-mindedness.[10]

By enlisting readers and viewers in his archival project, Lomax hoped to alter their day-to-day experience of human movement, new scientific understandings demolishing old prejudices. No longer would the "shuffling" movements of African Americans connote laziness; instead, they would tell a story about climatic adaptation, agricultural technologies, and dogged persistence. Indeed, as Choreometrics-trained observers moved through a city, they would encounter hard evidence about the long course of human history and human diversity in the body of every person they passed. A trip to the grocery store might teach as much as an afternoon at a natural history museum. Each moment would be like "looking through a microscope or underwater for the first time."

"All authors," Lomax wrote, "have their dreams. Mine runs this way: a folk dancer, an aboriginal choreographer, a student from some place away from the overwhelming mainstream picks up this book, looks through this atlas for his or her culture area and finds a pattern that is quite familiar—coming from his home or at least from his home ground." Though, until this point, he may have been suffocated by the barrage of Western cultural media, "now he discovers that there are many other aesthetic alternatives created far away from the urban art and pop marketplaces, including one by his own ancestors. This style he can feel in his joints and muscles belongs to him or is akin



Choreometric Flow Chart. FROM THE ALAN LOMAX COLLECTION AT THE AMERICAN FOLKLIFE CENTER, LIBRARY OF CONGRESS. USED COURTESY OF THE ASSOCIATION FOR CULTURAL EQUITY.

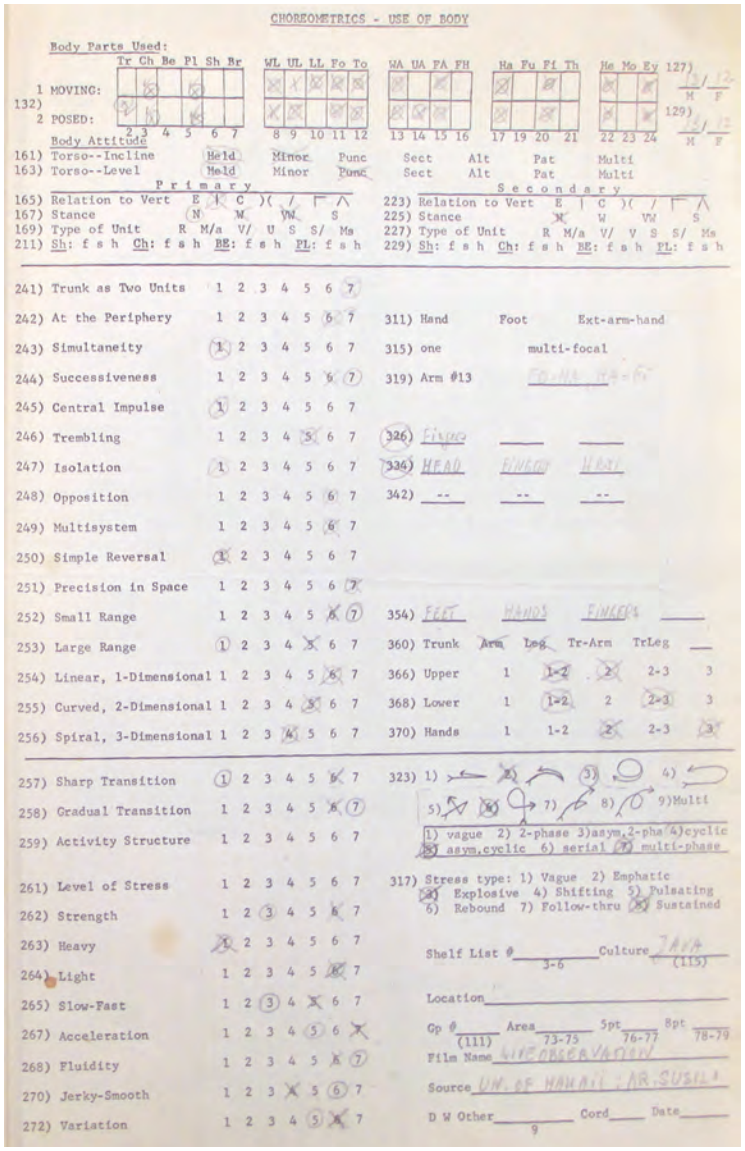
to the one he knows." He begins to "look at himself and his people with renewed esteem and begins to think, if he is a dancer, about what he can do with what he really knows. He has discovered that his own movement style is there, that it is composed of a special and fitting rearrangement of the same elements found in all human activity, but handled in an original style."[11] The natural body would return, triumphant, one more.

This was Lomax's holy grail. Numbers and figures, maps and diagrams, would awaken in even the casual

9 Paulay et al., "The Choreometric Coding System."

10 For more on similar midcentury efforts to inculcate open-mindedness and diversity, see Turner (2013) and Cohen-Cole (2014).

11 Alan Lomax, "An Atlas of Dance Styles," Box 4/18-04, The Alan Lomax Collection at the American Folklife Center, The Library of Congress.


CHOREOMETRICS - USE OF BODY

Choreometric coding sheets. FROM THE ALAN LOMAX COLLECTION AT THE AMERICAN FOLKLIFE CENTER, LIBRARY OF CONGRESS. USED COURTESY OF THE ASSOCIATION FOR CULTURAL EQUITY.

reader a new sense of culture, of history, and even of his or her own body. "Members of all the varied human traditions, whether they be viewers or program makers, film-makers or film goers" could "no longer easily be shamed or enticed out of their birthright."[12]

Unfortunately, Lomax's ideals fell short of his achievements. Crushed under the weight of its own ambition—and more than a thousand pages of data, photographs, and coding sheets—Lomax's book was never published. And while his films were broadcast on American public

12 Lomax, *Dance and Human Culture.*

Sevillana dancers in town square. FROM THE ALAN LOMAX COLLECTION AT THE AMERICAN FOLKLIFE CENTER, LIBRARY OF CONGRESS. USED COURTESY OF THE ASSOCIATION FOR CULTURAL EQUITY.

television, it is unlikely that they ever made it back to their communities of origin. Though Lomax saw Choreometrics as an interactive medium, a cybernetic system in a world of unidirectional broadcast TV, the final product ended up far more conventional.

With a characteristically modern idealism, Lomax believed in the power of vast accumulations to remake the world. In the end, however, it was not this transformational vision that determined Choreometrics' fate, but rather the far more mundane realities of budgets, page limits, and distribution contracts. Thus, just as much as Lomax's lofty idealism suggests that modern-day data-gathering projects might hold more radical potential than we have given them credit for, it is also a cautionary tale. In a world of seemingly frictionless exchange, it is easy to forget how profoundly dependent such projects are on the existence of robust technological, political, and social infrastructures. No matter one's intent, ideals without materials do not utopias create. ■

WHITNEY LAEMMLI *is a Postdoctoral Scholar at the Columbia University Society of Fellows in the Humanities.*

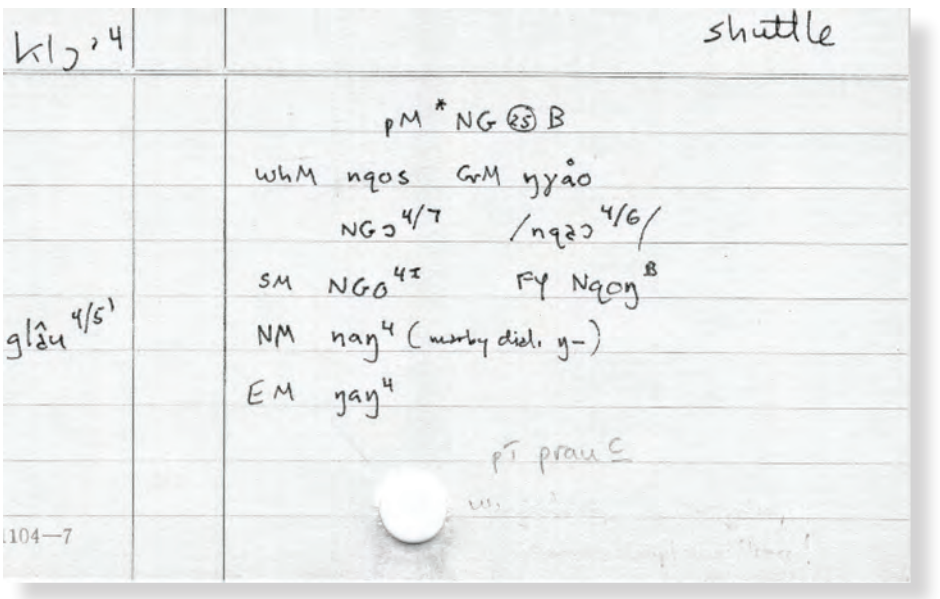
BIBLIOGRAPHY

Cohen-Cole, Jamie. 2014. *The Open Mind: Cold War Politics and the Sciences of Human Nature.* Chicago, IL: University of Chicago Press.

Szwed, John. 2011. *Alan Lomax: The Man Who Recorded the World.* New York: Penguin Books.

Turner, Fred. 2013. *The Democratic Surround: Multimedia and American Liberalism from World War II to the Psychedelic Sixties.* Chicago, IL: The University of Chicago Press.

Archiving descriptive language data



JUDITH KAPLAN EXPLORES THE POSSIBILITY OF A NEW GOLD STANDARD FOR ARCHIVING THE WORLD'S ENDANGERED LANGUAGE DATA.

TWO HANDSOME MAHOGANY BOXES, LONG AND NAR-row, sit on a high shelf in a home basement that is prone to flooding in the American Midwest. Release one of their tiny silver latches, and inside you will find a collection of vocabulary cards annotated in an impeccable hand. These cards constitute the sum total of known field data on Biao Min, a language of the Hmong-Mien family spoken by some 21,000 people in southern China. Compiled during a four-month research trip to Guangxi Province in 1982, they were filed away in a closet and forgotten for more than 10 years. Only in 2001 did these notecards come to the attention of the broader research community. Since that time, linguists and data curators have used them alongside other similarly vulnerable materials in a demonstration project called E-MELD that is designed to create both a repository and an infrastructure for the management of past, present, and future language data.

As this contingent and layered history suggests, problems of data management in linguistics are not new. They extend back at least as far as the early twentieth century, when a well-known fieldwork imperative took hold in American anthropology. Pushing beyond what had until then been a narrow focus on Indo-European language and culture, Franz Boas, Edward Sapir, and their students set out into the field—first with phonographs, then with tape recorders—to capture, transcribe, analyze, and ideally revitalize a host of Indigenous American languages (Darnell 2001). This work was dedicated to future generations of researchers and speakers alike. The resulting collections were preserved in text and audio formats by institutions like the American Philosophical Society and Indiana University's Archive of the Languages of the World. It was a race to fix the characteristics of thousands of languages before they changed beyond recognition or disappeared entirely (often at the hands of government acculturation programs). Speed and efficiency in the field were prioritized over any kind of long-term or systematic archival strategy (Swadesh 1954). Such work became a cornerstone of graduate education, a focal point of government programming, and a rallying cause for some speech communities. It gave rise to numerous rich, though unruly, collections of language data. Many of these, the Boas Collection not least of all, are undergoing rapid digitization today.

World War II put the brakes on this early boom in descriptive linguistics: prominent researchers left fieldwork in progress to join the war effort. At the 1944 meeting of the Linguistic Society of America, 80 out of a total 96 members in attendance reported that they were involved in "military crucial work" (Martin-Nielsen 2010). Defense funding flowed into the discipline, which rapidly gained institutional prominence and moved away from its roots in anthropology. This coincided with a shift in theoretical emphasis, from historical particularities to linguistic universals—a move that continued through the postwar period (Harris 1993). But concern with the loss of linguistic diversity—not unlike contemporaneous trends in the biological realm—by no means disappeared. By the 1990s, linguists were visibly in the field again, raising the profile of documentary and data-driven research within the discipline as a whole.

Marking this development, in 1992 the Linguistic Society of America (LSA) formed a Committee on Endangered Languages and their Preservation, which issued its policy statement on "The Need for the Documentation of Linguistic Diversity" shortly thereafter. This statement reflected the spirit of the day, justifying its recommendations via the benefits inductively to be won for "the study of universal grammar and linguistic typology." Expressing a level of disciplinary self-confidence that would have been unthinkable in Boas' day, the Committee intervened "for the sake of the future of linguistics, with the intent of enriching and preserving" the field. Specifically, it called upon academic departments to "support the documentation and analysis of the full diversity of the languages which survive in the world today," giving highest priority to those facing extinction and/or featuring highly divergent characteristics. Significantly, Committee members further recommended that data be "systematically preserved in a network of repositories which also regulate the availability of this documentation."[1] Such work was incentivized through the conferral of graduate degrees, hiring, promotion, and tenure priorities. It was reinforced over time by a number of organizations including the Endangered Language Fund, UNESCO, the Foundation for Endangered Languages, the Indigenous Language Institute, Terralingua, the Resource Network for Linguistic Diversity, the DOBES Archive, the Rosetta Project, and the Hans Rausing Endangered Languages Project.

Almost a quarter-century after the LSA first

1 http://www.linguisticsociety.org/sites/default/files/lsa-stmt-documentation-linguistic-diversity.pdf

published their recommendations, linguists and data curators are trying to wrangle the collections born of the last 100 years of "salvage" linguistics into some kind of order. The goal of projects like E-MELD and the Open Language Archives Community is just that: "to aid in the development of infrastructure for linguistic archives" (E-MELD 2000). For the architects of E-MELD, the mission is to address two serious problems facing documentary linguistics today: the rapid loss of linguistic diversity (current opinion estimates that roughly half of the languages spoken in the world today will disappear by the end of the century) and the rapid proliferation of independent digitization initiatives.[2] Governing here boils down to the cultivation of "best practices" for intermedial translation, and the development of metadata linking heterogeneous resources and concepts to one another. Furthermore, reinforcing a logic of collective and distributed effort in the digital preservation of language data, it extends to the relations among researchers who are expected to share resources and custodial responsibility. Such interoperability ideally holds out the promise for direct communication—across individual languages, technological platforms, and research traditions—without leveling linguistic diversity.

While E-MELD primarily addresses the needs of stakeholders in endangered languages research, in practice the project is as much about the protection of endangered archival materials: those two mahogany boxes. There is the sense that these can be revitalized through digitization. Ten case studies are featured in the project's "school of best practices," which is available to researchers around the world through LINGUIST List, the discipline's central online forum. Here, the project explores the nuts and bolts of moving between various media and a universally accessible web archive; the challenge being to move literally "From Notecards to the Web," "Shoebox to the Web," "Filemaker Data to the Web," prior integrative efforts like "TASX to the Web," and audio recordings on "Cassette to the Web." In the case of Biao Min, the task was to standardize digitization of the notecards—which maintain a window onto the comparative history of the Hmong-Mien family, an impression of the cultural life of Biao Min speakers, and vital characteristics of the language itself—in such a way that the resulting images would hold up for long-term preservation. Project members made choices about archival image format, user interface, data entry, and storage on the basis of this model collection that were meant to be generalizable. Moreover, they also applied themselves to the creation of resource metadata that would make the language intelligible within the framework of a hoped-for total linguistic archive.

The problems E-MELD faces derive from the logic of distributed responsibility—the need to host online archives, which can be quite sizable, at various sites—and the notion that data only have value when they can be found. What common infrastructure will allow linguists to identify relevant resources for a given study across archives? For example, what if a language of interest goes by different classifications or names in different collections (e.g., Lappish vs. Sami)? What if different structural tags are used by linguists in different traditions (e.g., possessive vs. genitive)? What if different systems of presentation are used (e.g., chronological vs. frequency-based vs. alphabetical)? And what if the resources themselves are submitted in formats that are wholly incommensurable (e.g., incompatible software tools; textual vs. recorded vs. video samples)? These are the kinds of questions motivating the search for a new total governing infrastructure.

Metadata, in this case, can be of two types: those that pertain to language *resources*, or to the languages *themselves*. The latter category is of rich conceptual interest because it blends top-down (theoretical) and bottom-up (descriptive) commitments about the characteristics of natural human language. Rather than hierarchically imposing a

2 http://www.unesco.org/new/en/culture/themes/endangered-languages/

BELOW: The UNESCO Atlas of Endangered Languages — Arctic Circumpolar.

new annotation system on top of the many already in existence—these being more or less suited to the particularities of the individual languages described—linguists and data curators have developed an "ontology" capable of linking extant, and possibly future, strategies for language analysis. This flexible and decentralized governing strategy has facilitated the recognition of new collective kinds: new groups of languages can now be compared, and therefore defined; new communities of researchers can interact and share their data; and new assemblages of archival objects can be brought together under the big tent of information.

Central to all this is the 2010 General Ontology of Linguistic Description (GOLD), which provides a formalized account of the most basic categories and relations used in linguistic description (GOLD 2010; Farrar and Langendoen 2003). With its roots in Scott Farrar's 2003 doctoral dissertation, GOLD allows linguists to search and compare *within* relevant resources (once these have been identified) using a standardized search vocabulary. To take an easy example, if a linguist wanted to look comprehensively within a corpus of glossed texts for examples of past-tense morphemes, he or she could invoke the GOLD term *PastTense* in a query, taming a babel of alternatives used in other linguistic markup schemes (e.g., *Past, PST, RemotePast, HodiernalPast*, and so on). A reduced ontology—one with ultimate compatibility with the Semantic Web—thus enables more languages, resources, and linguists to come together in a streamlined comparative framework.

The 100-year history I have just flown over reveals the emergence of a new *disciplinary* collective, one that is being defined—beyond Indo-European studies, Americanist anthropology, or the endangered languages community—by the web-based archiving of language data. Linguistics has been characterized as a field that depends on "second sourcing" its data: borrowing is widely accepted, as linguists understand that language learning and fieldwork are too labor intensive to be replicated continually from scratch (Lewis, Farrar, and Langendoen 2006). Thus, linguistics is cumulative, cooperative, and conservative with respect to data. E-MELD further illustrates how the desire to digitize and make web archives openly available has occasioned new methods of governance, ranging over increasingly general linguistic populations. But governing in this case has more to do with flexibility than control: taxonomy has been rejected in favor of ontology. With a radically simplified conceptual structure that articulates what are thought to be universal features of human language, recent adventures in linguistic data curation attempt to figure a new species-level population from the ground up. Whether or not these efforts will deliver a new gold standard remains to be seen. Historians and science studies scholars can ask in the meanwhile, what systems of value underpin contemporary efforts to archive endangered language data, and for whom do they apply?

JUDITH KAPLAN *is currently pursuing her interests in the history of the human and historical sciences (linguistics in particular) as a postdoctoral fellow at the Max Planck Institute for the History of Science, Berlin.*

BIBLIOGRAPHY

Darnell, Regna. 2001. *Invisible Genealogies: A History of Americanist Anthropology*. Lincoln, NE, and London: University of Nebraska Press.

E-MELD. 2000. Electronic Metastructure for Endangered Languages Data Grant Proposal. http://emeld.org/documents/E-MELD.html (accessed 3/10/15).

Farrar, Scott, and D. Terence Langendoen. 2003. "A Linguistic Ontology for the Semantic Web." *GLOT International* 7(3):97–100.

GOLD. 2010. *General Ontology for Linguistic Description (GOLD)*. Bloomington, IN: Department of Linguistics (The LINGUIST List), Indiana University. http://linguistics-ontology.org/

Harris, Randy Allen. 1993. *The Linguistics Wars*. New York: Oxford University Press.

Lewis, W., S. Farrar, and T. Langendoen. 2006. "Linguistics in the Internet Age: Tools and Fair Use." In *Proceedings of the EMELD '06 Workshop on Digital Language Documentation: Tools and Standards: The State of the Art*. http://emeld.org/workshop/2006/papers/lewis.pdf (accessed 4/12/15).

Martin-Nielsen, Janet. "'This War for Men's Minds': The Birth of a Human Science in Cold War America." *History of the Human Sciences* 25(5):131–155.

Swadesh, Morris. 1954. "Linguistic Time Depths of Prehistoric America: Penutian." Grant Report. Philadelphia, PA: American Philosophical Society.

Selections from the VALACO ARCHIVE

Vadig de Croehling, Director of Ideation, Process, and Interface at the Group for Research on Experimental Accumulation and Speculative Archives (REASArch), offers a sampling of elements from one of his organization's most inscrutable archival projects.

FOREWORD

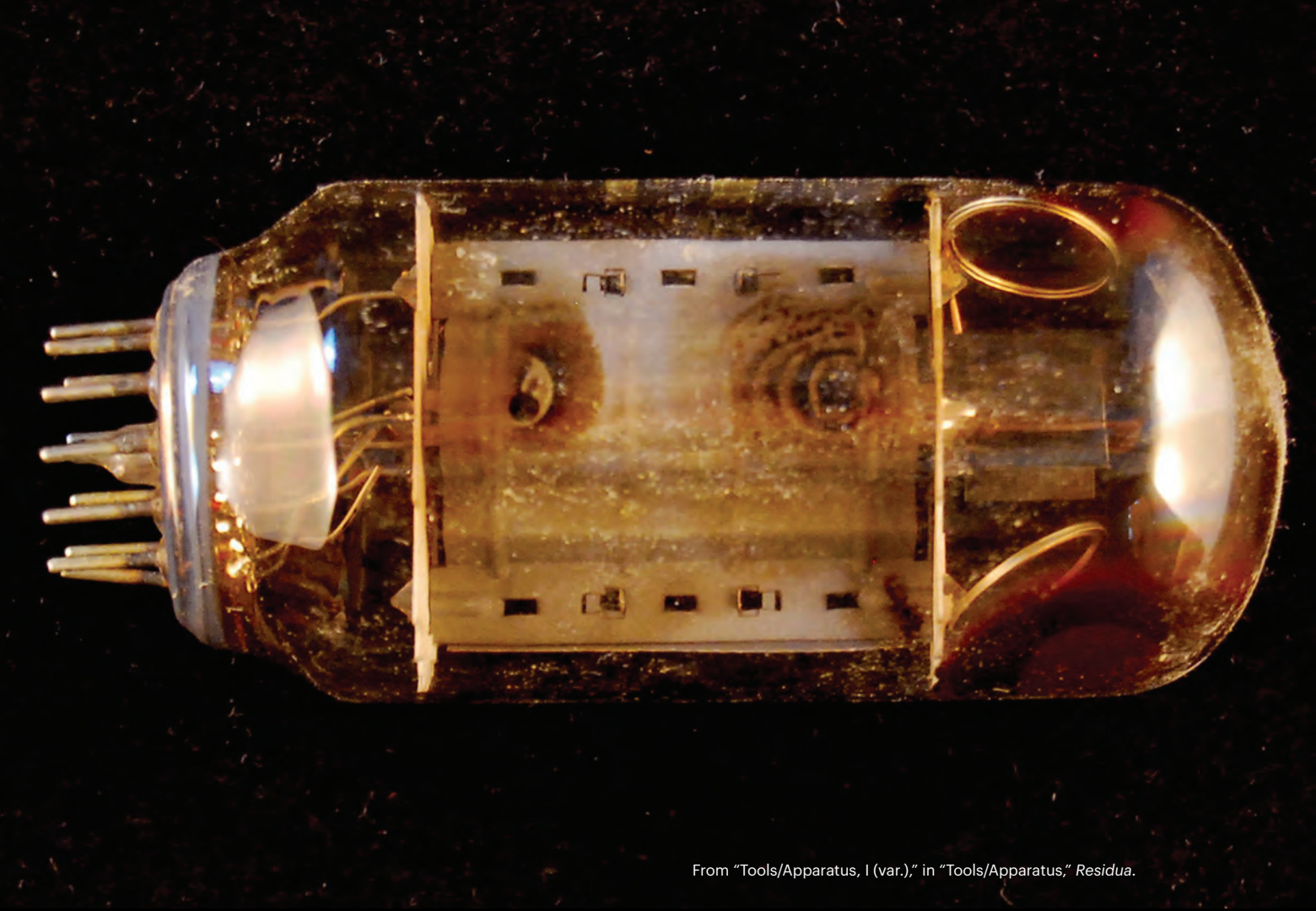
In the spring of 2004, a peculiar document surfaced in an obscure municipal archive in Buenos Aires, Argentina. Its contents—a motley assortment of idiosyncratic scrawls, organized into three unmarked manila folders—offered abundant evidence of a rare intellectual force, if few indications of the author's identity. One notation, etched in pencil inside the back cover of the third portfolio, furnished the following datum: "C. Roberto Valaco. *Schriftatlas*."

As subsequent inquiries have suggested, Roberto Valaco (born Robert Konstanz Wälke?) may have counted among some twenty-thousand German conscripts furloughed from the front lines near the end of World War II, in order to appear as movie extras in the epic costume drama, *Kolberg*, the Third Reich's last and costliest film production. Chimerical at best, Valaco's scant allusions to the experience in his "Atlas of Writing" would constitute the sole known testimony to date by any of the storied *Kolberg* extras.

Evidently a recluse, likely an autodidact, certainly a man of prodigious if enigmatic critical faculties, Valaco was consumed with issues ranging from the nature of memory and forgetting to the pursuits of scavenging and discarding, from the philosophical dimensions of vision to the vexed conditions of archival practice *per se*. Hence, for REASArch, the *Schriftatlas* offers but one unique inflection point—by turns seductive and inscrutable—about which to structure and continually restructure the Valaco Archive.

The materials provided here, a minute sampling of the larger project, will doubtless appeal to those concerned with questions of the *total archive*. We invite you to continue your researches at http://valacoarchive.com. Those with specific information relating to the life and thought of Valaco should contact Vadig de Croehling directly at vadigdecroehling@reasarch.com. Please stay abreast of the ongoing projects of REASArch at http://reasarch.com.

LEFT: From "Clichés VII (2nd Vertical/Horizontal Group)," in "Clichés," *Residua*.
ABOVE: From "Likenesses," I-IV, in "Démontage," *Forensics*.



From "Tools/Apparatus, I (var.)," in "Tools/Apparatus," *Residua*.

FROM "PRELIMINARY THESES ON THE NATURE OF THE ARCHIVE"*

- The archive is not the repository of its artifacts (though one may find artifacts in the archive); rather, the archive is a network of events both potential and conjectural.
- The archive does not so much contain, fix, or frame some thing or set of things (we may call these things the bodies of the archive) as it flashes up from the transitory, ineffable convergence of numerous errant bodies, the traces of which describe an ever-shifting frontier.
- The archive is no surrogate for cultural memory, nor is it the bedrock of an immanent, monolithic History (though memory and history alike may surely be summoned in and by the archive); rather, the archive is intrinsically multiple, constitutionally liminal.
- The archive does not equal, imitate, supplant, or otherwise eclipse some form of autonomous knowledge; rather, the archive is a horizon of such knowledge, just as any knowledge is but a horizon of innumerable archival impulses.
- One cannot maintain that "everything is an archive," just as one shall never alight upon an "archive of all things"; rather, one may discern, in any single thing, the rumor of untold archives past and future.
- The archive is not one and is no one. There is no one that is not, already, an archive beyond measure.

---

* In "Thesis/Metathesis," *Anarchaeolog*

EXCERPTS FROM THE SCHRIFTATLAS OF ROBERTO VALACO:

What of the collector impelled not by abundance and redundancy, but rather by poverty and concentration, by scarcity and singularity? What of the collection conceived as a projection of lost objects—that is, of things one can neither locate nor produce—of items to be sought and never found? What do we call that archive of collected objects dislodged a priori from the (linear) history of their acquisition? (I, 8a)

•

...Indeed, the image—and so the document, if not the datum in itself?—conceals within it the very diagram of its inevitable, innumerable reappropriations. That is to say, the image presages and prefigures something like a vast, indeterminate expanse of fertile if fallow pastures, to any one of which it may at some point be cast—as a spore by a subtle wind—and into any soil of which it may be sifted, sewn, hence reincorporated as though into a sward to which it seems proper, when in fact this sudden ground can be no less than fundamentally alien—hence, ontologically improper—to the ever wayward image.... An image, stated otherwise, is the harbinger of its own proliferation and diminution in space, a concentration of the improper, an infinite density of loss and the unending iteration or dispersion of that loss.... (II, 7a)

•

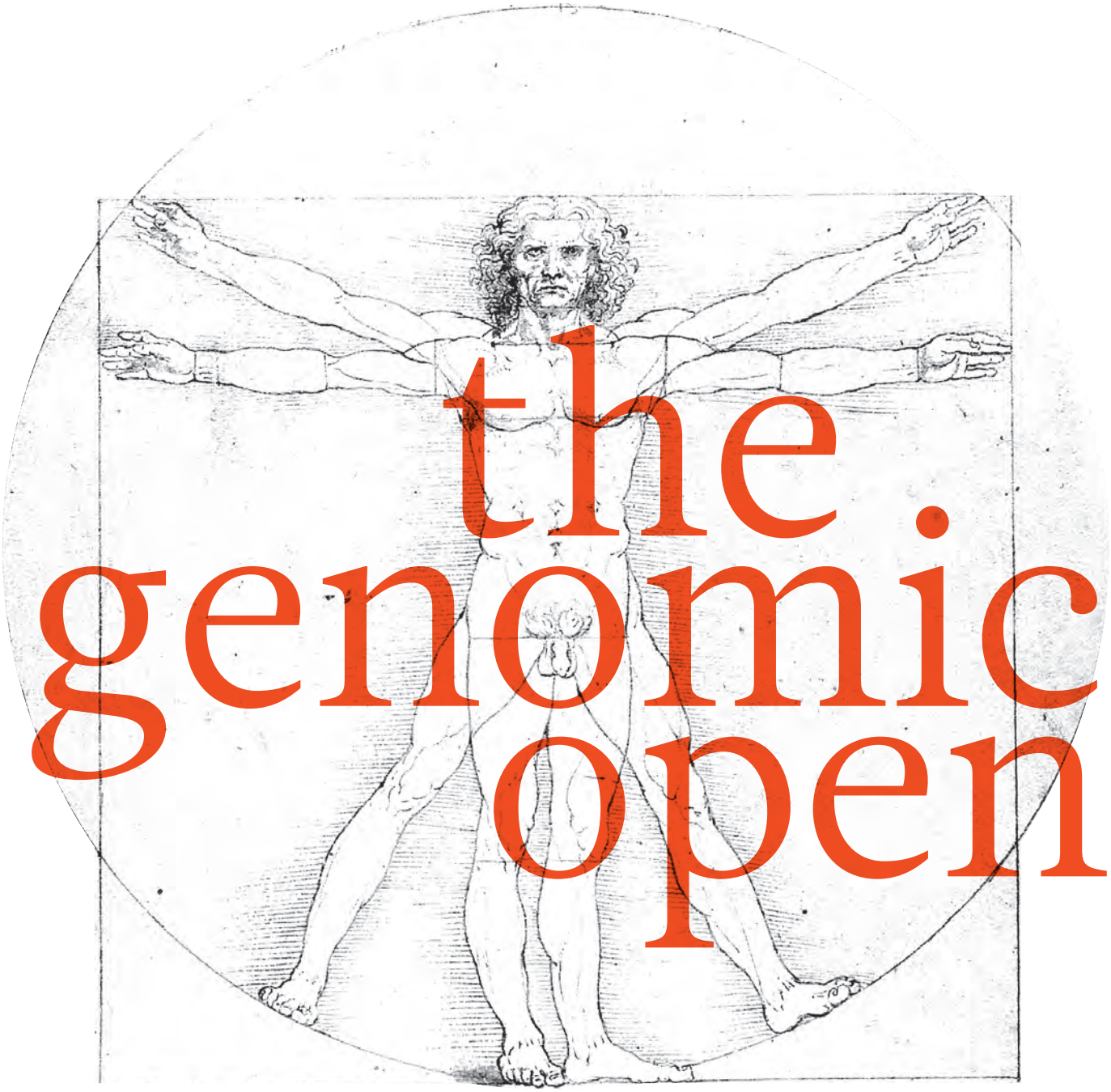
Madness is an image that resides between the absence of its object and the evidence of its past existence. Such an image mediates between the visible absence and the "improbable" evidence of a past materiality (some would call this a "truth"). This mediation—the medium—is, hence, a hallucination. I am inclined to think of it as a speculative archive.... (II, 11a)

•

...Memory as tumultuous ruin, or as shadow of thought, residue of vision? Memory as a damper on the friction of time.... Where there is but ruin, must we reinvent the foundation? Where there is nothing, are we doomed to invent its prehistory? (III, 8a)

VADIG DE CROEHLING *is an artist, curator, scholar of visual culture, and founding director of the Group for Research on Experimental Accumulation and Speculative Archives (REASArch). His work has appeared most recently (under the name of G. D. Cohen) at the Videoholica International Video Art Festival in Varna, Bulgaria; the 6th Cairo Video Festival, Medrar for Contemporary Art, Egypt; the Los Angeles Center for Digital Art (LACDA); and the Institute of Cultural Inquiry in Los Angeles. G. D. Cohen lectures on cinema and media studies and on Latin American visual culture at the University of California, Los Angeles, and serves as Associate Programmer for Los Angeles Filmforum, Southern California's longest-running screening series for experimental and alternative film and video.*

BELOW: "Twin Lens Reflex Triptych," from "Triptychs II (750)," in "Triptychs," *Anarchaeolog.*



the genomic open

Leaders of the Human Genome Project promised a genomic total archive. Jenny Reardon argues that their quest inspired visions of freedom and imprisonment vital to understanding today's ambivalences around open genomic data.

IF EVER THERE WERE A DOMAIN OF SCIENCE THAT DRAWS ON and fuels imaginaries of a total archive, genomics is it.[1] Genomics promises to reveal the secrets of life, to cure cancer, to solve the world's energy problems, but only if we create open access to *all* genomic data.[2] Life demands nothing less.

In the dominant popular account of genomics, scientists who pioneered genomic techniques while working on the Human Genome Project (HGP) fought an epic battle to create such an archive of genomic data that all could contribute to and access (Shreeve 2005; Sulston and Ferry 2002). In this story, a moral economy of science guided by the norms of openness and communalism struggled to survive in the face of the growing power and widening influence of a capitalist economy.[3] Judgment is clear: leaders of the public effort to sequence the human genome were heroes; Craig Venter and his venture capitalist backers were villains.[4]

In these accounts of the HGP, the quest to share human genome information—to create an open genomic archive—motivated scientists working on the public Human Genome Project, and led to their passionate commitment and personal sacrifices. At my own institution, the University of California, Santa Cruz, Jim Kent reportedly spent day and night in his garage writing code, stopping only to ice his wrists (Townsend 2015). Yet, questions arose on the ground about the value and meaning of an endeavor that required an ever-growing number of

1 The story of how genomics arose out of and fuels contemporary aspirations for a total archive is complex one for which I can give only a very partial account. There are many parts of the story that I cannot even begin to address: for example, how it came to be that anyone thought genomics might contain the complete information—the code—of life. For this, see Lily Kay's excellent historical account (2000).

2 And increasingly many other kinds of data too: environmental, health record, drug surveillance, social media, and other "omics" data, just to provide a short list. For a list of grand things that genomics promises to bring us, pick up any popular account of genomics. For example, see Collins (2006) and Venter (2007).

3 The term "moral economy" was coined by E. P. Thompson in the1960s to described the norms and practices that regulated exchange during the eighteenth-century bread riots. It was imported to the history of science by Robert Kohler to describe *Drosophila* geneticists' practices of exchange (see Kohler 1994; Thompson 1963).

4 While Venter receives most of the critical attention, Randy Scott, the Chief Science Officer at Incyte Genomics Inc., attempted to patent ESTs as early as 1991. Further, Scott never sought to make genomic data public, while Venter did (Venter 2007).

DNA-SEQUENCERS. The sequencing floor in BGI Hong Kong, showing the Illumina Hiseq 2000 sequencers. PHOTO BY SCOTTED400 – WIKIPEDIA. LICENSED UNDER CREATIVE COMMONS ATTRIBUTION 3.0 UNPORTED.

automated sequencers to displace humans and to consume large amounts of reagents and capital (Sulston and Ferry 2002:177). Who and what benefitted from these informatic and automatic infrastructures designed to create, store, and manage ever-expanding archives of genomic data? Despite the valiant effort to defend public science, many on the ground feared that genomics installed a technocratic and capitalist regime at the heart of the life sciences in which power inhered in the few who had the money to buy and operate sequencing machines.[5]

Their concerns did not go unstoried. Many genome scientists wrote accounts of the HGP that brought to the fore these deeper structural transformations that unsettled understandings of the value of shared data—total or otherwise—as the grounds of knowledge and the public good. Illustrative is John Sultson's *The Common Thread* (Sulston and Ferry 2002). Sulston led the UK arm of the HGP, and is widely recognized as a hero of the effort to defend public access to human genome sequence. Yet the book's opening line makes clear that his would be no simple tale of triumph: "I just heard the prison door close behind us" (Sulston and Ferry 2002:1).

This startling description recounts the moment Sulston realizes that he has signed onto the HGP, and there is no going back. In the pages that follow, he chronicles the transformation of his everyday work life as it moves from the small intimate spaces of intense human interaction at the Laboratory of Molecular Biology (LMB) in Cambridge to the vast cavernous spaces of the sequencing machines at the Sanger Center, built in Hinxton. We learn of a life no longer marked by late-night encounters in the lab, coffee time, drunken punting expeditions, and Guy Fawkes celebrations, but one lived under the pressure of keeping an army of sequencing machines running on schedule (Sulston and Ferry 2002:50). Sulston reluctantly, but seemingly inexorably, enters a capitalist world of production. The amounts of money required continually threaten to overreach the capacities of public governments and private foundations, making the HGP vulnerable to a venture capital takeover (Sulston and Ferry 2002:99). Sulston receives multiple offers to join private industry, offers he reportedly always took seriously. And while he accuses Venter of no longer being in science, but in business, just a few pages on in his account he explains that he too was in business: "Bob [Waterston] and I had

5 At the time, I heard one geneticist describe the situation as nothing short of a Maoist struggle of the people waged this time not with guns, but with sequencing machines.

the biggest businesses at the time; Eric Lander aspired to have the biggest business" (Sulston and Ferry 2002:189).

These businesses sought to operate at ever-greater speeds and efficiency. For this, they required ever-greater resources and tightly controlled management. Not everyone could take part. Indeed, many would be excluded. At the end of the HGP, of the 20 listed as authors on the *Nature* paper, only two, Sulston argues, had "the high level of industrial organization needed to accelerate the production of sequence": the Sanger Center and Bob Waterston's lab at the University of Washington in St. Louis (Sulston and Ferry 2002:203). Despite his allegiance to the ethos and practices of openness he attributed to the LMB, Sulston found himself a central character in this transformation of his field of biology into an industrial-scale production system that excluded all who could not keep up.[6] It was, along with other domains of knowledge, becoming a part of informatic capitalism.[7]

Knowledge, Francois Lyotard argued a decade before the launch of the HGP, had become "an informational commodity indispensable to productive power." Indeed, he asserted, it was "a major—perhaps *the* major—stake in the worldwide competition for power" (Lyotard 1979:5. Certainly, by the mid 1990s, actors central to genomics acted as if this were the case. Craig Venter and his financier Wally Steinberg justified their entrée into the race to sequence the human genome as nothing less than an effort to "save America's biotech industry."[8] By 2000, the fortunes of the U.S. stock market hinged on events in human genomics, and world leaders took an active role in its governance (Sulston and Ferry 2002:247).[9]

While predicted by a social theorist a decade prior, Sulston, a biologist, lived through and described these changes. "Biology," he observes, "had undergone an economic sea change—it now held the promise not only of tremendous knowledge and great benefits to humankind but also fabulous wealth. As biologists we had lost our innocence" (Sulston and Ferry 2002:209).[10] The changes left untouched all dimensions of the scientific life Sulston had known: its buildings, its practices, and even the core value of openness itself. By the end of the HGP, it was the very goal of an open and total archive of genomic data that fueled the goals of production and wealth accumulation that befit the sequencing machines. After Venter announced at the Cold Spring Harbor meetings that ABI was going to fund a company to sequence the human genome, both Wellcome and the National Institutes of Health (NIH) greatly increased the amount of funds they devoted to the HGP, entering the sequencing machines "arms race" (Sulston and Ferry 2002:220).[11] In the fall of 1998, The Sanger Center bought 30 of the new ABI capillary sequencing machines at $300,000 apiece; Eric Lander at the Broad Institute bought 125. In 1999, the year after Venter launched Celera, ABI sold a billion dollars' worth of sequencing machines. If there was any clear winner in the race to complete the human genome sequence, it was this manufacturer of the machines (affectionately known by some genome scientists as "Arrogant Beyond Imagination"; Shreeve 2005:60).

Sulston and others at the Sanger Center did try and resist this big-money dimension of genomics. Reportedly, Tim Hubbard, then head of sequence analysis at Sanger, explored the possibility of using a "copyleft" agreement developed by the free software movement to protect the public project's human genome data (Love and Hubbard 2005). Such an agreement would have provided a formal legal meaning to the HGP's principle of open access, specifying that all were free to use HGP genome data but could place no restrictions (e.g., patents) on its further development. However, those who oversaw the public genome databases reportedly strongly objected. The data, they argued, should remain free for all to use in whatever way they saw fit, including patenting and licensing further development and redistribution of the data (Sulston and Ferry 2002:238; see also Cukier 2003).

There was indeed no going back, not even for openness. The power to sequence—and thus to play a major role in the genomics revolution—was already concentrated in a few institutions. Wealth had become a major stake in biology. Inequalities between researchers—both among the genome mappers and sequencers and subfields of biology—became institutionalized as the price paid for universal access to the sequence of the human genome.[12]

Today, the injunction to share and to make available all genomic information— not just of the human genome, but all genomes—is driven once again by a belief in the

6 On the centrality of speed to the HGP, see Fortun (1998).

7 For a definition of informatic capitalism as I use it, see Franklin (2012). Key components are the rise of information as a dominant commodity form, the simultaneous and entwined production of markets and informatics and their re-working of both labor and knowledge.

8 Steinberg argued that NIH did not have the resources to compete with Japan, Britain, and Germany (Venter 2007:158).

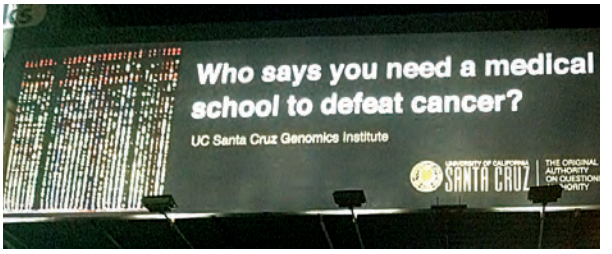
9 In March of 2000, UK Prime Minister Tony Blair and U.S. President Bill Clinton made a public statement affirming the Bermuda principle that genomic data would be made freely available. Immediately, the Nasdaq—the index of high-technology stocks—lost 200 points. Biotechnology stocks lost $30 billion in value in one day alone. The value returned when the President's science advisor and Francis Collins clarified that the principle of openness did not prevent use of the data by private companies.

10 Note that Sulston is referring here to only a small band of biologists: those involved in genome sequencing. The extent to which they represented the leading edge of broader transformations that led to all biologists "losing their innocence" is still a matter of debate among historians, sociologists, and anthropologists of science (Sulston and Ferry 2002:209).

11 How scientists were able to persuade governments to support public genomics in the late 1990s—a time marked by privatization of formerly public sectors—deserves further explanation. One key to this puzzle is the way in which publicness easily figured as a form of openness that fostered efficiency in informatic work. In other words, as Sulston would later make clear, openness was not just a political goal; it was a practical and technical one. For a further exploration of the alignment of "public" genomics with an informatic/technocratic conception of openness, see Reardon (forthcoming).

tremendous power of genomics. Consider this billboard I encountered as I was riding my bike home from a movie in January of 2015.

The billboard stands at a busy intersection in San Francisco where tens of thousands pass by every day. These messages about the power of genomics are joined by messages about the importance of sharing one's DNA and data. Perhaps later, while on Facebook, Bay Area citizens will come across MeForYou.org and learn how they can help another UC, the University of California, San Francisco (UCSF), through sharing their DNA and medical records (http://meforyou.org). UCSF launched the MeForYou.org social media campaign to "put a new thought into the public consciousness" that sharing genomic data and medical records not only helps science, but also communities and loved ones. MeForYou, UCSF spokesperson David Arrington argues, is part of an effort to create a new "social contract" with biomedicine in which people agree to share their data in exchange for new knowledge that helps all people (Lu 2013). The moral force of the initiative is strong. Who, after all, would not want to help Georgia, the young girl who is the MeForYou poster child? Dissent—or even public dialogue—under these conditions is difficult.



GENOMICS DEFEATS CANCER. San Francisco billboard. PHOTO BY AUTHOR

These dreams of total access to all genomic information inspire and create visions of new routes to universal knowledge and democracy. Yet these commitments to the open flow of genomic information exclude and exert control in ways that led even genomics' most ardent supporter to invoke the imagery of a prison. We are in need of languages and frameworks that allow us to grasp and speak about these powerful and paradoxical dimensions of our ever-deepening commitments to total archives in an age of bioinformatics.

JENNY REARDON *is Professor of Sociology and Director of the Science and Justice Research Center at UCSC, and is author of* Race to the Finish: Identity and Governance in an Age of Genomics *(Princeton University Press, 2005)*

12 Shorett, Rabinow and Billings captured these changes at the time in their 2003 commentary in *Nature Biotechnology* (2003). Sulston is explicit throughout his account that genomics created inequalities among researchers. In the end, those with the most efficient sequencing operations captured most of the funding. The norm of openness itself favored those with more resources who would not be hurt by making data public before articles were published, and who had the labor and money to upload data to GenBank (see Reardon forthcoming).

BIBLIOGRAPHY

Collins, Francis. 2006. *The Language of God: A Scientists Presents Evidence for Belief* New York, NY: Free Press.

Cukier, Kenneth Neil. 2003. "Open Source Biotech: Can a Non-Proprietary Approach to Intellectual Property Work in the Life Sciences?" *The Acumen Journal of Life Sciences* 1(3). http://www.cukier.com/writings/opensourcebiotech.html

Fortun, Michael A. 1998. "The Human Genome Project and the Acceleration of Biotechnology." In *Private Science: Biotechnology and the Rise of the Molecular Sciences*, edited by A. Thackray, pp. 182–201. Philadelphia: University of Pennsylvania Press.

Franklin, Seb. 2012. "Virality, Informatics, and Critique; or, Can There Be Such a Thing as Radical Computation?" *Women's Studies Quarterly* 40(1/2):153–170.

Kay, Lily. 2000. *Who Wrote the Book of Life? A History of the Genetic Code*. Palo Alto, CA: Stanford University Press.

Kohler, Robert. 1994. *Lords of the Fly: Drosophila Genetics and the Experimental Life*. Chicago and London: University of Chicago Press.

Love, James, and Tim Hubbard. 2005. "Paying for Public Goods." In *Collaborative Ownership and the Digital Economy*, edited by R. A. Ghosh, pp. 207–229 Cambridge, MA and London, England: MIT Press.

Lyotard, Jean-Francois. 1979. *The Postmodern Condition: A Report on Knowledge*. Minneapolis: University of Minnesota Press.

Reardon, Jenny. Forthcoming *The Postgenomic Condition: Ethics, Justice, Knowledge After the Genome*. Chicago, IL: University of Chicago Press.

Shorett, Peter, Paul Rabinow, and Paul R. Billings. 2003. "The Changing Norms of the Life Sciences." *Nature Biotechnology* 21(2):123–125.

Shreeve, James. 2005. *The Genome War: How Craig Venter Tried to Capture the Code of Life and Save the World*. New York: Ballantine Books.

Sulston, John, and Georgina Ferry. 2002. *The Common Thread: A Story of Science, Politics, Ethics, and the Human Genome*. Washington, DC: John Henry Press.

Lu, Stacy. 2013. "Want to Help Heal the World? Start by Sharing your Health Data." *TedMed*, May 13. http://blog.tedmed.com/tag/meforyou/

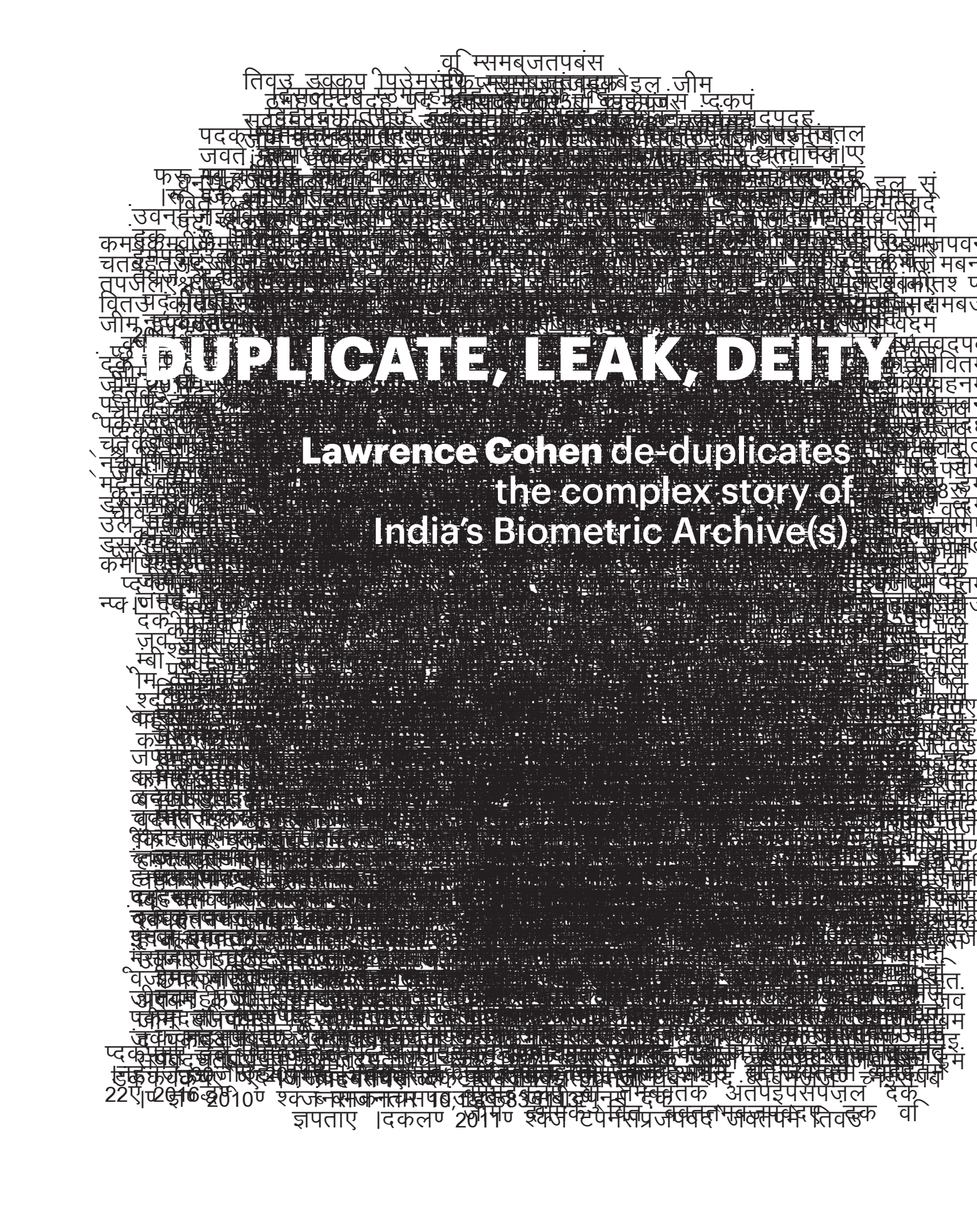
Thompson, E. P. 1963. *The Making of the the English Working Class*. London: Victor Gollancz.

Townsend, Peggy. 2015. "A Visionary, a Genius, and the Human Genome." UC Santa Cruz Genomics Institute, June 22. https://genomics.soe.ucsc.edu/news/article/152

Venter, J. Craig. 2007. *A Life Decoded: My Genome, My Life*. New York, NY: Viking.

DUPLICATE, LEAK, DEITY

Lawrence Cohen de-duplicates the complex story of India's Biometric Archive(s).



Q: How would you describe the visualisation scene in India?

A: It would grow because of the same reasons worldwide, the sheer amount of data is growing rapidly…. I was walking in one of the schools and saw [a] huge pile of students' annual report card dump[ed], being a data junkie my heart sunk at seeing all valuable student data rot away silently.

Interview with Report Bee CEO Ananth Mani (Kirk, 2011)

IN SEPTEMBER 2015, INDIAN PRIME MINISTER NARENDRA Modi—once banned from the United States for his apparent role in orchestrating anti-Muslim pogroms in 2002—returned stateside, traveling to Silicon Valley to promote a vast flotilla of e-governance initiatives called Digital India. At its core is what has been widely termed the "JAM Trinity": J for Jan Dhan Yojana, promising bank accounts to the poor; A for Aadhaar, the national biometric program promising to "de-duplicate" all duplicitous claims on state services in cash or kind; and M for mobile phones, the vehicle enabling the new "cashless society" JAM promises.

Digital India was rolled out just after Modi's first year in office. It appeared to centralize digital government, which for the past decade had been split at the national level between at least two ambitious programs, the National Population Register (NPR), tied to border security, and the Unique Identification Authority (UIDAI), with its "Aadhaar" ID form (*aadhaar* means "basis" or "foundation"). Each program promised to collect the biometrics of all Indian subjects, a process known as *capture*.

Digital India is under the purview of the Department of Electronics and Information Technology, or DeitY. The godly acronym existed before the 2014 election and is not an invention of Hindu right-wing ideologues within Modi's Bharatiya Janata Party (BJP). For both secular and religious identified blocs across parties, Digital India illustrated the emerging promise and debatable hubris of a new technocracy claiming self-consciously superhuman, panoptic powers. The infotech pantheon was *henotheistic*,, in the sense claimed for Hinduism by the nineteenth-century Indologist F. Max Müller of a single deity uniting multiple divinities (1878). One may worship the goddess Siva, Vishnu, or myriad other valued divinities, Müller suggested, but one worships each as the One. Louis Dumont would later define such a relation between values as *encompassment* (1981).

In the digital pantheon over the past decade, the main divinities were UIDAI and NPR. Each promised a national archive of biometric governance that would identify all Indians: voluntarily for UIDAI, and by law for NPR. Each was building an archive to digitize traces of all persons in India, and each claimed the primacy of its archive against the other as the proper form and substance of a new kind of collective entity, what we might call *nation-as-archive*. Digital India and its JAM comprise an explicitly henotheistic mode of governance, encompassing both of these emergent, overlapping, and often competing biometric archives as a single political form, one closely identified with the PM and his charismatic authority.[1]

If the population and its nation were mobilized as a visceral collective in the consolidation of European urban, colonial, and settler modernities through the emergence of statistical devices and the conception of a model, the nation-as-archive emerges as something else. We might turn to current historicizations of machine-learning approaches to big data by their architects—of big data constituting an emergent condition of plenitude organized less around statistical modeling than around data storage, curating, and algorithms enabling "visualization"—to convey a sense that the collective form at stake is an unrelentingly expanding mass of data in itself, a *different figure of mass* than that of the mass body and one that demands new conditions of governance. The point is not that such whiggish historicizations of big data (e.g., Gray 2009) are adequate to a history of reason or the archive, but that they offer a feel for the contemporary, for a widely available sense of collectives and their government as not only *dependent on* an immensity of information (the familiar ground of a biopolitics), but ontologically *constituted as* information.

The opening epigraph, from a boutique collection of "data visualisation stories from around the world," gestures toward a collective form, one in which data—like organic matter—"rots," in which the relationship between organic matter and data undergoes some kind of material-semiotic shift (Kirk 2011). The care of the child is here organized less around the rotting of sequestered or poorly distributed food stockpiles than around the rotting of piles of information. It is not only that such data is "dark," in the sense of not yet monetized, but that its life

1 Digital India "cuts across multiple Ministries and Departments" and "weaves together a large number of ideas and thoughts into a single, comprehensive vision so that each of them can be implemented as part of a larger goal. Each individual element stands on its own, but is also part of the larger picture. Digital India is to be implemented by the entire Government with overall coordination being done by the Department of Electronics and Information Technology (DeitY). Digital India aims to provide the much needed thrust to the nine pillars of growth areas…" (DeitY 2015). It might be taken as pandering to left critique to note the requisite phallic language (thrusting pillars). But my provisional reading would be that such language mobilizes the foundational figure of the pillar and in effect links the imaginary of one program (UIDAI)—organized around an airborne and motile vision of platforms flexibly bearing the weight of the state and of the nation's biological need—to that of another program (NPR), organized around more conventional, grounded metaphors of the sovereign control of territory.

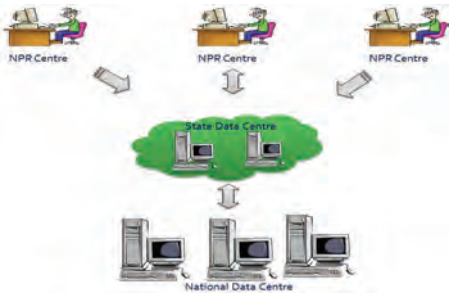


FIGURE 1. NPR Pyramid from Census of India. "The NPR Process".

FIGURE 2. National Database.

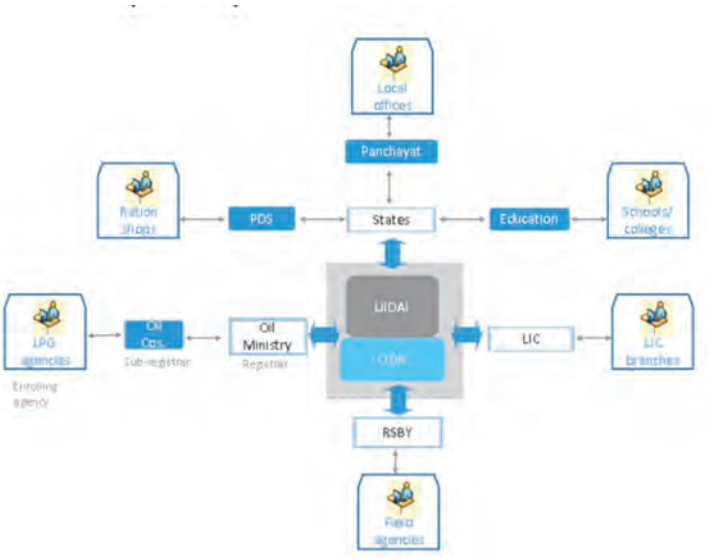


FIGURE 3. "The Unique ID Agencies" from UIDAI Strategy Overview, April 2010.

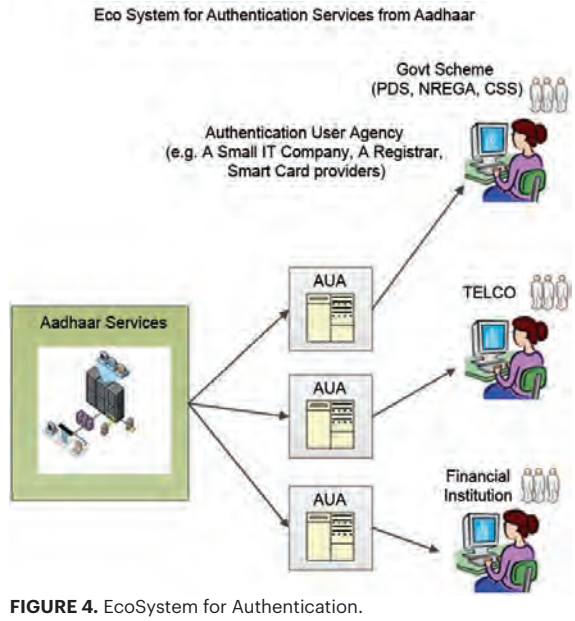


FIGURE 4. EcoSystem for Authentication.

festers or degrades.[2] Nation-as-archive similarly gestures toward an emergent terrain in which the nation is a database and governance depends on the care of its archive as a kind of living thing. This terrain involves a host of newly mobilized things: the silo and its loneliness; the loss and recovery of the social; security and its proliferating rationalities; and the transfer of "service" or "benefits" and the governmental problem of distribution.

In the first the decade of the 2000s, both NPR and UIDAI found different paths toward capturing the biometrics and variable amounts of biographical data of as many residents of India as possible, and each entity vied henotheistically to encompass the other's archive. NPR's conception of archive was centralized, as Figure 1 suggests.

As its architects imagined, the "National Data Centre" extended and intensified the idea of a strong central government, here an inverted pyramid in which information appears to fall into a single repository. Identified with the passionate attachment to the singular nation and with a sedentarist, hyperterritorial conception of those comprising it, this gravitational archive was, in Benedict Anderson's, conceptualization, a bound seriality (1998: 29–45). It has been repeatedly represented as a central place or thing, collecting a wide range of territorial, demographic, and biographical information. National strength correlates with the quantity of information: multiple data fields for biography and territory fall together into one.

UIDAI's public presentations of privacy protection similarly address its "highly secure data vault," variably identified as its "Central ID Data Repository," or CIDR. These presentations intensify in response to civil society criticism of UIDAI and its Aadhaar that centers on privacy. But Aadhaar's architects, in contrast, stress how *little* information UIDAI collects into the CIDR, and how this minimalist archive is more relevant as a platform (a more useful translation of *aadhaar*) that links together myriad "silos" of information, forming an "ecology" or "federation."

When pyramids do appear in UIDAI's self-representation, these are often turned on their side; it is less a repository through the sovereign force of gravity than a catalytic enabler of a range of goods. The box in Figure 4 labeled "Aadhaar services" places the secure central archive as part of an already distributed ecology charged with redistribution.

NPR and Aadhaar invert the relation of citizen and resident in different ways. What would become NPR began after the 1999 Kargil war with Pakistan as an effort to create a biometric identity database to distinguish Indian citizens in Kashmir from presumptive infiltrators. Biometrics carried future promise and were linked

2 This rendering of data as organic—or conversely, this digitization of rot—reprises the familiar racialized organicism of the (post)colony as garbage (Anderson 2010; Chakrabarty 1991; Kaviraj 1997).

to multiple biographical data to stress a proper relation to space: border security would be effected by linking the collection of a hyperterritorializing plenitude to the promise of indelible physical traces. This conception of archive was progressively scaled up over a decade, from the Indian Muslim to the Indian citizen, from Kashmir to the nation.

But how to achieve this larger scale? Bureaucrats and contracted experts associated with the Interior Ministry proposed piggybacking NPR on the Census of India. The Census was not an archive of citizens, but of *residents*: whoever was enumerable across the terrain of the nation. Using the Census's preexisting infrastructure to achieve the needed archival scale meant that *residence* and not *citizenship* became the condition of biometric subjectivity. The focus on internal security specified the unit of biometric data collection as the citizen, with biometrics offering the promise of distinguishing that citizen from its double or "duplicate": the fake citizen or terrorist.

NPR was never closely associated at the executive level with the emerging constellation of information technology capital and expertise. It drew not only upon the preexisting infrastructure of the Census, but also upon preexisting standards of administration, hierarchy, and contract in the creation of new governmental forms. Like other administrative units, it was subject to the familiar accusation of corruptibility, not only at the level of bureaucratic procedure but within the constitution of the digital archive.

By contrast, UIDAI organized itself around corruptibility as a problem. It moved away from standard governance—viewing corruptibility as requiring human solutions—to corruptibility as a machine engineering problem, one of databases, not bureaucracies. It is commonly narrated as being the brainchild of Nandan Nilekani, a founder and the CEO of the IT outsourcing giant Infosys, a company that pioneered a range of identity instruments to organize and credentialize IT service labor. Like many nouveau hyper-rich, Nilekani was troubled by the persistence of a massive and "leaky" state bureaucracy and its cozy relation to a small coterie of elite family capitalists, a situation preventing the efficient management of poverty and weakening entrepreneurialism. Nilekani offered a blueprint for completing the country's neoliberal transformation in his 2008 bestseller *Imagining India* and was brought into the previous Congress Party-led government to create Aadhaar.

Nilekani's concept in brief is that India's future increasingly depends upon the distribution of "service," principally forms of welfare in kind or, increasingly, cash, but corruption "leaks" out a significant proportion of this wealth, both through rent-seeking by petty bureaucrats and other office-holders charged with service distribution and through the production of "duplicates," fake or copied identities in the list of persons or households entitled to a service. The conception of service is organized around a biopolitical figure of bare life: of *residents* within or moving across a terrain who must be supplemented by services to survive and to thrive. Aadhaar's early critics from the political right worried that its basis only in residency (as opposed to citizenship) would enable undocumented Bangladeshi migrants to gain official status and receive undeserved state services by getting Aadhaar numbers.

UIDAI's own concern with wastage was not the unsubstantiated specter of the migrant, but the general corruption or "leakage" of legitimate claims on distribution by most persons. Archives, and in particular databases, are rendered efficient and governable through consistent "de-duplication," ensuring that all items in the collection are "unique" and thus curtailing leakage.

De-duplication is a technical term that addresses problems of storage efficiency, of record variability and the need for correction, and of security from duplicate (e.g., stolen) identifying objects.

Efficiency: "de-duplication is a task of identifying record replicas in a data repository that refer to the same real world entity or object and systematically substitutes the reference pointers for the redundant blocks; also known as storage capacity optimization" (Faritha Banu and Chandrasekar 2012:364).

Correction: "data sources are independent… [adopting] potentially inconsistent conventions" (Maddodi et al 2010:664), so to build an effective "data warehouse," data "has to be transformed and cleaned before it is loaded into the warehouse" (Chaudhuri et al. 2006). Data may differ across source archives because of different schemas by which they were formed, and thus cleaning involves "schema extraction and translation" (Thakare et al. 2015:10). Data difference may not only involve the cultural difference of distinct schemas, but also the problems introduced into any given source archive by human error, which constitute "dirty data" (Maddodi et al. 2010:664). The distinction between what makes data untranslatable, requiring schema extraction, and what specifically renders it dirty is not always clear in this literature. Archives, as products of assemblage, appear to present translation as both a semiotic and arguably a moral problem.

Security: the presence of duplicates in an archive when each of those duplicates refers to the same object (say a given resident of India) and provides a means for different users of the archive to make different claims as or for that object, as, for example, when the hero or villain in a movie gains access to the nuclear arsenal through a duplicated identity.

In creating UIDAI, Nandan Nilekani argued that for India to become more like China, a developing economy powerhouse, it needed to be de-duplicated *as a nation*. Neoliberal efficiency, the security of the commonweal in the face of mass corruption, and the translation problem of what we might term *history-as-assemblage*, were all gathered up into a single technocratic repertoire. Corruption was rendered as a matter of either duplication from above, the large-scale seeding of an archive with duplicates inserted by powerful interests exercising control over it, or duplication from below, the fake identities upon which persons—urban migrants, slum dwellers, landless laborers—unrecognizable within the formal ar-

chive may depend.

If duplication from above depends on control of archival infrastructure, UIDAI proposed a radically new and independent archive. The problem for its engineers was the social itself, the network of interest and biographical relations that limit fair and efficient distribution and produce leaks. They proposed collecting as little biographical and locational information about persons as possible—assurance would depend on biometrics and not biography—to produce a deterritorialized archive cut off from the duplicative nature of the biographical and social. The subject of this archive was a body offering ten fingers and two eyes, officially a "resident": incorruptible and free from political tampering because the Aadhaar numbers issued to all residents of India would convey no information, no history. Each time this resident sought a service, the plan presumed, he or she would present a body part and the system would return a "yes" or "no": you are you, or you are not you. As more and more persons were signed up, and as more and more services were linked, India would be de-duplicated.

The subject of this archive was a body offering ten fingers and two eyes, officially a "resident": incorruptible and free from political tampering because the Aadhaar numbers issued to all residents of India would convey no information, no history.

De-duplication did not require UIDAI: multiple parallel and derivative data-cleaning projects emerged around the same time. The customer list of a small cooking gas distribution agency near Delhi whose owner and manager I know was considerably reduced when every gas customer had to present proof both of identity and of residence to the agency, which was then turned over to state auditors. Whether or not people presented their Aadhaar numbers or other forms of legitimate ID as proof, the exercise de-duplicated the list by more than half. The "corruption" of households availing themselves of multiple subsidized gas cylinders was curtailed, as was that of gas deliverymen siphoning off small amounts of gas or police pressuring gas deliverymen for a cut.

Nilekani's dream to remake India demanded de-duplication of service distribution at a massive scale. The archive had to scale up to the nation. Unlike NPR, the Census was inadequate for UIDAI to produce an identity archive—to produce India—at such a scale. Rather, public-private partnerships (PPPs) were set up in most Indian states to be independent from current bureaucracy: subcontractors were paid per new biometric registrant, profit expansion and not national infrastructure drove scale, and the server network was designed to test and retest subcontractor skill and honesty.

As Aadhaar became both a promise of inclusion for those too marginal to have access to earlier modes of identity, and a threat of Big Brother as the universal platform making life through service possible, it began to appear the very condition of citizenship. Civil society activists on the left argued that UIDAI would not just link the "silos" of individual service distribution programs through its platform, but produce a condition of total *convergence*. Despite UIDAI's insistence on an ecology of multiple silos federated through its Aadhaar network (as opposed to an NPR-like National Data Centre), and its claim that it archived almost no personal information besides a registrant's biometrics, the drive by its engineers to make Aadhaar the platform for any "service" from food subsidy to credit suggested that any form of value in belonging to the nation would need to come through Aadhaar. This was a new kind of citizenship: UIDAI lacked any statutory right under Indian law to mandate its Aadhaar identity, to serve as the necessary platform of service delivery, or to be the vehicle of de-duplication.

Some UIDAI engineers that I interviewed formally responded to their progressive critics that the UIDAI archive, unlike NPR, eschewed biography, did not in itself collect information on Aadhaar registrants, and would maintain a federation of silos, not the feared convergence. Privately, however, some UIDAI engineers told me that if politicians wanted to use Aadhaar to converge silos, they could. When I reported this internal concern to one of the most senior UIDAI engineers, he got upset: we have created a corruption-free identity, he said. But if politicians and social forces misuse it, there is a limit to what we can do.

Nilekani and his team fought to prevent the powerful senior officials aligned with the Interior Ministry and NPR from getting control of UIDAI. For UIDAI, the state security apparatuses—including NPR and other central repositories of identity—were each silos that could be more effectively governed if Aadhaar became their universal platform. For the NPR team, UIDAI was simply a different and parallel way to gather data, and if it promised efficiency, it was yet another contracted tool for national information to be encompassed by the demands of the National Data Centre.

Digital India's publicity in 2015 offered an explicitly flexible account of information and its government. Existing bureaucratic structures across the range of state service were expected to open themselves to new norms of easy access, no longer dependent upon the power of the bureaucratic office and of its rent seeking. Existing archives of identity could be flexibly deployed to manage and audit this access. Concerns with both *physical* leakage—the wrong people on the wrong side of the border—and also with economic leakage—the proliferation of duplicates wasting the commonweal—were to be secured through the interrelation of what we might call *neo-Aadhaar* and the Modi persona itself.

Under the previous, Congress Party-dominated central government, Nilekani and his allies in the Indian Planning Commission—the dominant mandarinate of the development state—envisioned UIDAI's success on the model of other PPPs free from the bureaucratic ("social") entanglements of lesser arms of the state. UIDAI was set

up in relation to the Planning Commission and the Finance Ministry, but was largely autonomous from them. UIDAI evaded the party politics of the parliamentary system and was not constituted as a statutory body. But as Aadhaar increasingly came to be constitutive of a new form of citizenship, its critics launched a series of court cases challenging its legality. In 2014 and 2015, the Supreme Court of India affirmed that no service could require that people register for an Aadhaar card.

There were other challenges. Before Narendra Modi won the 2014 election, Aadhaar's fate seemed politically as well as constitutionally unclear: the program was closely identified with Congress President Sonia Gandhi, and Nilekani himself, despite his frequent disavowal of social and political corruption, had been pressed to run for office. He, like Congress, lost.

But Modi, victorious, would go on to embrace Aadhaar with a vengeance. News accounts and popular stories began to circulate about the new PM's panoptic ability to know what was going on in all senior political and bureaucratic offices, and that he was having Aadhaar scanning devices placed in every major government office to ensure that officials were present and that their output could be measured. Aadhaar, with its reputation under Nilekani for placing the nation-as-archive outside of and protected from the bureaucratic office—that is, the conventional institutions of the state—was being brought in to manage those very institutions. If Aadhaar had been designed to disentangle office from service, it was now synonymous with a new government of office. Beginning in late 2014, I heard an emergent class of panoptic Modi joke in which an official skipping office duty, breaking a rule, or not following the PM's instructions would suddenly get a phone call from Modi himself.

Beginning in July 2015, Digital India loosened this close connection between the panoptic Modi and Aadhaar. Whether or not Aadhaar itself would be the primary identity archive for the new e-governance seemed less important, particularly given its questionable legal future, than did its formal apparatus: biometrics, ever more universal scanners, and some kind of henotheistically constituted lattice of future identity archives serving as the platform layer for the state and for finance. In July, I heard stories of a "secret" pact between Modi and Nilekani to keep Aadhaar's powerful linkage of the nation's silos intact. Over the next months, UIDAI and its Aadhaar program were placed within JAM, a commitment to shift all service to direct cash transfer via the explicit trinity of universal bank accounts, Aadhaar biometric scanning to ensure de-duplication, and mobile phones as the sites across which the anticipated regime of microcredit and microspending would be enacted.

It is clear is that the division between a centrist and state-based national archive (the National Data Centre of NPR) and an exceptional nation-as-archive located across a vast federation of silos managing welfare, health, education, credit, labor, and so forth (Aadhaar as universal platform beyond the reaches of state corruption) no longer seems to hold. Modi as panopticon may have diminished somewhat, but the controversial leader's image and persona girding a new ethic of state office has been linked to Nilekani's promise of a guarantee of identity and service based on the separation of service and office. ■

LAWRENCE COHEN *is Sarah Kailath Professor of India Studies in the Department of Anthropology at the University of California, Berkeley.*

BIBLIOGRAPHY

Anderson, Benedict Richard O'Gorman. 1998. *The Specter of Comparisons: Nationalism, Southeast Asia, and the World.* London: Verso.

Anderson, Warwick. 2010. "Crap on the Map, or Postcolonial Waste." *Postcolonial Studies* 13(2):169–178.

Chakrabarty, Dipesh. 1991. "Open Space/Public Space: Garbage, Modernity and India." *South Asia* 14(1):15–31.

Chaudhuri, Surajit, Ganti, Venkatesh, and Kaushik, Raghav. 2006. "Data Debugger: An Operator-Centric Approach for Data Quality Solutions." *Data Engineering* 29(2):60–66.

Department of Electronics and Information Technology (DeitY). 2015. "How Digital India will be realized: Pillars of Digital India," http://www.digitalindia.gov.in/content/programme-pillars (Accessed Feb 22, 2016).

Dumont, Louis. 1981. *Homo Hierarchicus: The Caste System and Its Implications.* Chicago: University of Chicago Press.

Faritha Banu, A., and Chandrasekar, C. 2012. "A Survey on Deduplication Methods." *International Journal of Computer Trends and Technology* 31(3):364–369.

Gray, Jim. 2009. "Jim Gray on eScience: A Transformed Scientific Method." *The Fourth Paradigm: Data-Intensive Scientific Discovery*, edited by Tony Hey, Stewart Tansley, and Kristin Tolle, pp. xvii–xxxi. Redmond, WA: Microsoft Research.

Kaviraj, Sudipta. 1997. Filth and the Public Sphere: Concepts and Practices about Space in Calcutta. *Public Culture* 10(1):83–113.

Kirk, Andy. 2011. "Data Visualization Stories from Indiea, By Ananth Mani (Interview)," *Visualizing Data,* August 30th, 2011. http://www.visualisingdata.com/2011/08/data-visualisation-stories-from-india-by-ananth-mani/ (Visited Feb 22, 2016).

Maddodi, S., Attigeri, G. V., and Karunakar, A. K. 2010. "Data Deduplication Techniques and Analysis." *Emerging Trends in Engineering and Technology, 3rd International Conference*, pp. 664–668. New York: Institute of Electrical and Electronics Engineers.

Müller, F. Max. 1878. *Lectures on the Origin and Growth of Religion: As Illustrated by the Religions of India.* London: Longmans, Green and Co.

Nilekani, Nandan. 2008. *Imagining India: Ideas for the New Century.* New Delhi: Penguin Books.

Thakare, Manisha R., Mohod, S.W., and Thakare, A.N. 2015. "Various Data-Mining Techniques for Big Data." *IJCA Proceedings on International Conference on Quality Upgradation in Engineering, Science and Technology* 8:9–13.

Finn Brunton goes inside the Bitcoin blockchain to explore the weirdly meticulous collective archive, and how it might someday govern us.

KEEPING THE BOOKS

"First, the Company was forced to assume all public power. (The unification was necessary because of the vastness and complexity of the new operations.)"
—Jorge Luis Borges, "The Lottery in Babylon"

MANY OF BORGES'S STORIES FOLLOW A SIMILAR ARC: some seemingly small, innocuous thing—an encyclopedia, a lottery, the act of dreaming or trying to write a novel—expands in scope and scale until it becomes indistinguishable from its context, like his famous 1:1-scale map that completely covers its territory. In "The Lottery in Babylon," the administration of a lottery grows to incorporate misfortunes as well as winnings of all kinds, and extends to every citizen, until the operation of the lottery effectively becomes the state (and, latterly, something more like fate itself). Part of Borges's sly joke in this story is that the seemingly cruel and arbitrary actions of the Company in charge of the lottery are actually *preferable*, as a mode of governance, to those by which people are still elevated and ruined largely by chance: a chance skewed, rendered impure, by wealth and power. There's a deeply seductive appeal to governance by an inhuman system: however byzantine the nested layers of the lottery become, there's a random draw at the center of it that can't be bribed, intimidated, or begged for mercy. What makes this system "inhuman," given that there are few activities more human than staking an outcome on the turn of a card, and that every step of the rewards and punishments expresses our all-too-human convictions? Can you call a lottery a government? How could you defer authority to a system you *know* has nothing at the center, nothing but pure chance? We shake our heads together in puzzlement.

Welcome to Bitcoin.

Or, rather, welcome to the "blockchain," the system that underlies Bitcoin. Like Borges's lottery (that most wasteful of civic activities) that becomes the state, Bitcoin is a largely experimental, novel form of currency—an idea somewhere between "visionary ambition" and "kooky absurdity"—whose underlying mechanism, the blockchain, is being transformed into the technological substrate for a new, abstract kind of governance. The blockchain is a payment system with no money; a single, canonical record that is copied everywhere and maintained by everyone; a quasi-system of government whose ultimate authority rests on a series of deliberately useless, arbitrary computational problems. This state isn't built completely around a lottery, but rather around a *ledger*.

We start with physical cash to understand how this ledger, the blockchain, works, because they share a common problem, one that's far more challenging to address with digital cash: making unique objects that are easy to produce and difficult or impossible to reproduce. When I hold currency—let's put a U.S. $20 bill on the table now—I have an object with a very particular set of constraints. It must be almost exactly like every other U.S. $20 issued by the Treasury so it can function as legitimate money. But it must also be unique: if there is a single other bill *exactly* like it, one of them is a counterfeit. The bill must be very easy and cheap for the Mint (and a small set of textile and printing organizations) to produce, and yet nearly impossible for any other group to reproduce. There is no other bill like this one before us (serial number JB9557548B, 2009 series, Timothy Geithner's signature, a little ballpoint pen squiggle over the portico of the White House), but there are 6.4 billion others that are very, very close.

Meanwhile, the history of computing and telecommunications is primarily the work of transmitting perfect copies over imperfect channels, whether those copies are in the RAM and the hard disk of a single computer, or on a screen and a server on different continents. It is not enough to say that digital objects can be copied (with the connotation of a degraded, knockoff version): they can be *duplicated*, by design, thanks to decades of brilliant research devoted to reliably producing and verifying bit-for-bit duplicates of files.

Unique objects, yet perfect duplicates. You can already hear the grinding friction between the words "digital cash." The "cash" part is crucial; individuals can transact cash directly without having to pass through a "trusted third party"—a credit card payment or an online payment from our bank account. The besetting problem of digital cash research and development throughout the last two decades has been to produce a digital object that could be easily generated, transmitted, recognized, and exchanged—but not duplicated—without relying on a third party like a central bank, a clearinghouse, or the state. We should be able to transact this "cash" without creating new money objects or new copies of existing money objects. The Bitcoin blockchain's answer to this seemingly intractable problem of digital objects acting as money: don't have objects.

There's no string of characters that constitutes a bitcoin, no file or set of bits or bitcoin "thing." All that exists are *addresses* in the ledger, which represent bitcoin *ownership*; bitcoins don't exist apart from their attachment to an address. Think of it as an archive that has rich and meticulous documentation of provenance and chains of custody without any actual documents or artifacts. It resolves the complex legal and technical distinctions between data and metadata, text and paratext, by having *only* metadata. These transactional records and ownership logs constitute the existence of "bitcoins."

All the exchanges of ownership between Bitcoin

PREVIOUS PAGE:
20 July 1917, Secretary of War Newton D. Baker, blindfolded, drew the first draft number in the lottery to be called up: Number 258. "US LOTTERY, 1917" U.S. NATIONAL ARCHIVES' LOCAL IDENTIFIER:165-WW-420(P379) FROM: AMERICAN UNOFFICIAL COLLECTION OF WORLD WAR I PHOTOGRAPHS, COMPILED 1917 - 1918 (RECORD GROUP 165)

addresses are broadcast on the network; these transactions are settled, or confirmed, every 10 minutes. *Settlement* means that everyone running the Bitcoin protocol software—all the peers on the peer-to-peer network—takes the latest transactions on the system and races to solve a cryptographic problem that will link the "block" of new transactions with the previous blocks, which in turn are linked into the chain. The problem is difficult enough that most of the community would have to work together to post false transactions, double-spend money, or otherwise mess with the system. The winner of the solution race gets some new bitcoins, in the form of new records of ownership that didn't exist before. In other words, what makes new money in this system—what the money is, in a literal sense, made of—is the record of the existence and circulation of the money thus far. (The solutions to the problems are meaningless, exceedingly improbable results of slowly escalating difficulty to keep the rate of settlement and the production of new money constant.)

This is, therefore, an "append-only public ledger." It is a record of events—transactions between addresses—that everyone maintains (public) and to which new events can be added but not removed or altered (append-only). As of this writing, the ledger held 77,219,785 transactions. At first, the ledger was stored mostly on personal computers and custom-built servers in backyard sheds and basements; now it is kept in massive installations in cold regions of the world with inexpensive electricity and high-bandwidth Internet connections. It's nearly 20 gigabytes in size, and not just from transactions.

THE LOTTERY EXPANDS, WRITES BORGES, from merely contributing to the vicissitudes of human life to apportioning power: "I have been proconsul," says his narrator, and "I have been a slave. I have known omnipotence, ignominy, imprisonment." Very quickly, blockchain users and developers realized that an append-only public ledger—a system, collectively maintained, that only confirms that an event took place at one time, never to be changed, edited, or denied—could serve as a kind of archive, and then as the bare-bones foundation of a contractual order that could create companies, even minimal governments. The collective maintenance meant that, seen in a certain light, the blockchain was a robust, distributed archival backup system. If you could incorporate something into your transaction, it would be swiftly stored on hard drives all over the world; thus, the blockchain now includes 2.5 megabytes of diplomatic cables from WikiLeaks, a thousand digits of pi, texts from the *Bhagavad Gita* and the Pope, ASCII art and Valentine's Day messages, and encoded images and mysterious encrypted files.

This archival property of the ledger—complete with timestamps and planet-scale redundancy—also made it ideal for the sorts of activities previously relegated to notaries, such as witnessing contracts. More than ideal, in fact, because the blockchain could be used as the basis of automated contracts that could publicly document their own fulfillment, and could even accrue and arrange payment out of the blockchain itself. Carefully designed blockchain contracts could become the basis for "decentralized autonomous organizations" (DAOs), institutions that operate largely without human guidance and share out rewards to human "employees" for their contributions. DAOs connected together, requesting work and distributing resources, have been proposed as a system of experimental, minimal government written in scripting language, the libertarian dream realized of society assembled out of contractual relationships. The conditional is important here; some of these advances *could* happen, and several organizations are rapidly building on the blockchain—whether Bitcoin's or their own, comparable version—to make them viable, most notably the "smart contract" platform Ethereum.

Before they get into the mire of practice, before the messy dissolution of the first blockchain-based marriage (marriage vows are a favorite hypothetical test case for smart contract architectures), or the tangles of offshore "autonomous" operations dodging taxes, storing files, and making payments in currency units like satoshis, szabos, dogecoins, and litecoins, we can see the conceptual implications of the blockchain with greater clarity. The blockchain's simple, abstract promise is to trust neither in people nor the state, but in a set of cryptographic properties. Given those properties, a system of records can be perfectly and publicly maintained. Borges imagined a society that mitigates the injustice of the human condition by submitting *everything* to rigorous, inhuman chance, to the total lottery. The social imaginary in the blockchain is still stranger: that money, contracts, even law and government, can be built on nothing but meticulous, automated, collective maintenance of the archive. ■

FINN BRUNTON *(finnb.net) is an assistant professor in Media, Culture, and Communication at NYU.*

The blockchain now includes 2.5 megabytes of diplomatic cables from WikiLeaks, a thousand digits of pi, texts from the *Bhagavad Gita* and the Pope, ASCII art and Valentine's Day messages ...

7070B49DBF3BA12BEA427CB6651ECF7860FDC3792268031B77711D63A8610F4¶CDA551B7FB331103889A62E2CB23C0I

EXHIBIT: THE ENTROPY

WHAT DOES A PERFECTLY RANDOM ARCHIVE LOOK LIKE? FINN BRUNTON EXPLAINS.

F85362BBA49B9E0086D1DA0830E4389AB1 9C14511F0B355D4E8964A058ADA6EFB6A3FDD225E9E102EB6C163D23DF

ARCHIVES

THE NATIONAL INSTITUTE FOR STANDARDS AND Technology (NIST) maintains an archive of the generic and the default called the Standard Reference Materials. It is an archive of a very specialized sort: not one of particular objects, but instead of the most generic objects possible, against which others can be benchmarked, simultaneously completely typical and as specific as exacting measurement and engineering can make them. NIST has cigarettes for testing the ignition resistance of furniture, waterway sediment, crude oil, slurried spinach, argillaceous limestone, and reference peanut butter. Starting on September 5, 2013, a bit before noon, they also began producing standard random objects at a rate of one per minute: strings of 512 bits of entropy, broadcast every 60 seconds. (They call it a "public randomness service.") The first one starts like this: "17070B49D ...".

This "public randomness beacon" starts with the combination of two independent pieces of hardware that generate random numbers. The resulting 512-bit number is an excellent source of randomness, which is then combined with all the data pertinent to that particular value: the version number, the timestamp of its creation, output frequency, a status code, and—most significant, for the question of archives—the value of the *previous* output, the most recent random broadcast. This collection of data is then "hashed," or run through a function that takes data of any length and produces data of fixed length so that any change to the original data changes the hash output. You dump the data in the hopper, and get a string of characters ("63C4B71D51...") that preserves the original randomness while also being trivial to verify that it corresponds to its time and status information. This string is then signed with NIST's private key, a cryptographic tool for proving that NIST in fact sent it; that collection of data is hashed again, and at last you have the output value.

The result is an abstract kind of archive with a set of powerful properties, a set of characters that contains no information in a mathematical sense—each character is unpredictable based on past activity, and tells you nothing about the next character to come—but that can also verify that it is the product of NIST, sent by those who claimed to send it.

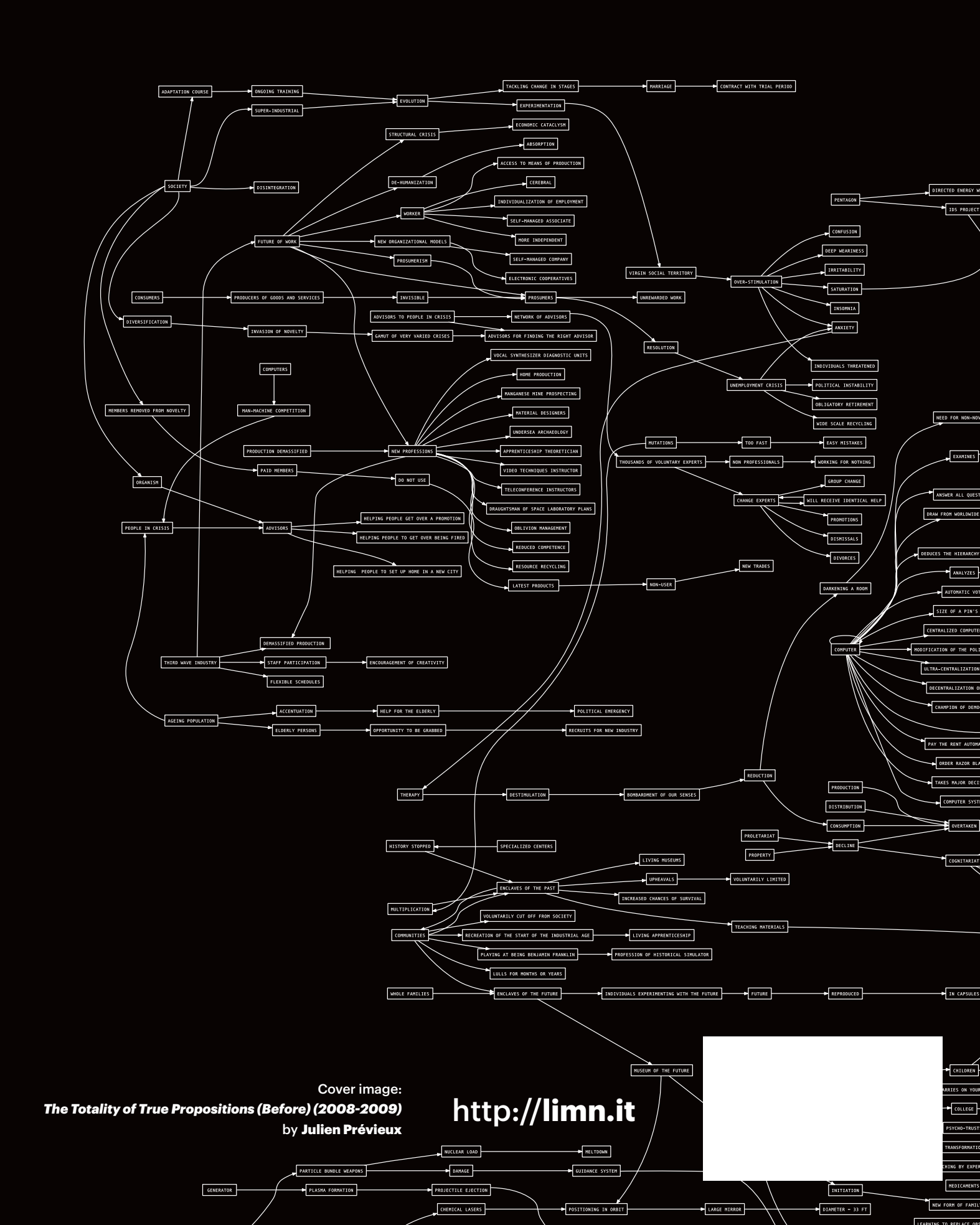
Let's say you need to randomly recount ballots from some districts to verify the integrity of a vote. How can everyone be sure you're using actually random numbers to choose the districts? If you get to select the numbers, you could rig the election. So you use the output of the public randomness beacon. What if you and your co-conspirators fake the output of the beacon? Maybe you can generate fake "random" characters and get access to NIST's private key to sign them, so you know in advance what the random draw will be. Now think about the reliance on random quality assurance checks in manufacturing everything from cars to pharmaceuticals, in conducting medical screenings, as components of stock market trading strategies, and even in military decisions—the safest evasive maneuver, all other things being equal, is one your opponent can't predict because even you don't know it in advance—and the importance of having *reliable* unreliable numbers becomes clear.

This is why the *archive* of entropy is so vital: because every new random string is hashed with the previous one, it is easy to verify the output of a hash, and the output is extremely difficult to predict in advance. This means you can generate a fake "random" string, and steal NIST's key to send it out, but anyone can check whether it incorporates the randomness of the previous broadcast. Which it won't, unless you faked that one too, but that in turn needs to incorporate the broadcast before it, link by link, two years back, a minute at a time. In other words, to produce randomness everyone can trust—randomness that reveals no information about future randomness, a perfectly level probability landscape—it has to be part of an archive of a very special kind, a timeline that verifies nothing but its own integrity. ■

FINN BRUNTON *(finnb.net) is an assistant professor in Media, Culture, and Communication at NYU.*

https://beacon.nist.gov/home

PHOTO: PIERRESELIM



Cover image:

The Totality of True Propositions (Before) (2008-2009)

by **Julien Prévieux**

http://limn.it