"Data" is the new "Stuff"

Sophia Toubian

Information Studies 298A: Doctoral Seminar: Research Methods and Design

December 21, 2023

Contents

The proliferation of "big data" and the subsequent permeation of the term "data" into everyday language has fundamentally shaped our modern conceptualization of information. What constitutes data, its usages, and its perceived value underpin many facets of society from interpersonal communication to national policy making. Yet amidst the sociotechnical algorithms, intellectual debates, and broad definitional ambiguity, there persists a lack of nuanced vocabulary to describe specific categories and forms of information. This research confronts the shift to a ubiquitously singular and dominant framework that positions "data" as the catch-all descriptor for every conceivable manifestation of information.

Such terminological simplification obscures critical distinctions in how various individuals and institutions understand, interact with, and assign value to information across contexts. More than a merely semantic concern of word choice and connotations, the lenses through which information is labeled, defined, and bifurcated on a fundamental level inevitably construct the schemas by which it is conceptualized and leveraged for myriads of purposes, from quotidian decisions to sweeping data-driven initiatives. Without reconsidering existing terminology conventions, the current paradigm continuously inhibits comprehensive literacy and evaluation of information types ranging from qualitative personal experiences, to quantitative government statistics, to interpretivist humanistic analytics.

This research project aims to investigate potential alternatives to the reductive "data" designation through methodological exploration of more clearly delineated vocabularies for information types. Proposed studies encompass critical surveys, discourse analysis, experiments, ethnographies, and computational linguistics spanning public and academic spheres. Centering emerging critical data studies scholarship alongside subdisciplines of information science, the project integratively examines the genesis of existing terminology frameworks alongside the

implications of paradigm shifts toward new systematic information lexicons. Fundamentally, by evolving beyond the universal "data" vocabulary limitation into domain-specific, context-conscious information type classification systems, individuals and institutions may substantially transform understanding, valuation, usage, and conversations surrounding the ever-proliferating information that defines modern society's relationship to knowledge production.

Additionally, the increased prevalence of "big data" and data-driven decision-making has positioned quantitative information framed reductively as "data" at the forefront across sectors, further cementing the use of this limited vocabulary. This paradigm lacks sufficiently nuanced terminology to describe the multitude of forms, types, and contexts of information. As scholars contend, the very terminology used to conceptualize information shapes understandings and perceived value, constructing the paradigms of meaning-making (Drucker, 2011).

The core problem this research confronts is the shift to a ubiquitously singular "data" framework applied as the catch-all term to encapsulate information in all manifestations. Such oversimplified labeling limits literacies, obscures critical distinctions, and constrains comprehensive evaluation. This project's purpose is twofold: to critically assess the impacts of "data" terminology dominance on conceptualizations of information forms, and to explore the implementation of alternative category-conscious information lexicons. This entails methodological exploration of shifts away from reductionist "data" vocabulary limitations across disciplines.

Literature Review

History and Definitions of "Data"

In *The Emergence of Probability: A Philosophical Study of Early Ideas about Probability,*

Induction, and Statistical Inference (2006), Ian Hacking provides an extensively researched intellectual history focused on the origins of mathematical probability and induction concepts intertwined with statistical inference. By excavating early scholarship by figures like Pascal, Fermat, Huygens, Bernoulli, and others, Hacking exposes the gradual conceptualization of calculating chance and constructing quantification techniques still underlying modern usages.

One of Hacking's major focuses lies in disambiguating terminology like probability as a degree of belief versus relative frequency, the latter eventually turning into a basis for inferential statistics. Hacking closely tracks evolving definitional debates that seek to categorize chance situations based on subsets like hypothetical, empirical, and statistical syllogisms. His detailed excavation of attempts to systematize language for precise calculation resonates with this paper's terminological aims.

While mathematicians constructed probability grammars aiming for stability and universality, Hacking explores contextual influences like gambling problems or evaluating testimonies in court. By highlighting the continuity between ethical, legal, and financial domains and the emergence of statistical quantificational tools, Hacking provides an antecedent case study relevant to arguing terminology's co-constitutive shaping of data practices today.

Hacking's seminal work modeling the conceptual maturation of statistical probability through closely attending to categorical language negotiations offers methodological precedent for studying the conceptual and practical implications of terminology choices related to contemporary data cultures. Hacking's philosophical analysis ties together epistemic perspectives, quantification techniques, and real-world application contexts in a manner highly pertinent to this study's aims.

Alain Desrosières' *The Politics of Large Numbers: A History of Statistical Reasoning*

provides a historical analysis of how numerical information and statistics have been generated and used for political and social purposes over the past few centuries. Desrosières traces the evolution from the earliest quantitative state records to the emergence of national accounting systems aiming to statistically represent the economy and population.

A core argument is that the production of statistics is not a purely technical process, but inherently political - numbers can construct simplified representations to enable governance, shape social realities, and privilege certain worldviews or epistemologies. Therefore, behind statistics are always normative choices and conventions determining methodology, categories for classifying and counting, and framing effects.

For instance, Desrosières examines how censuses and national income accounting classifications Create categories like "employment" or "GDP" that structure perceptions but are not objectively pre-existing. The use of quantification and statistics grows alongside state administration, enabling centralized monitoring and intervention. Number formats also enable links across distance and abstraction, from individual phenomena to summaries.

Overall, Desrosières provides a theorization and genealogy of the interplay between state power, knowledge infrastructures, and quantification. He shows how political decision-making utilizes statistics to designate, classify, and intervene in population groups and economic flows. In turn, these numbers construct epistemological conventions about social facts. This interdependency has only intensified with contemporary datafication and usage in policy settings.

The key relevance to this project is the demonstration across history that numerical information and statistical categories carry epistemic conventions, theories of governing, and worldviews that shape their meaning, usage, and effects on society over time. The terminology

used is not neutral but rather situates numbers politically.

<u>Objects vs. Meaning</u>

Michael Buckland's 1997 paper works to reconceptualize "documents" to move beyond written text to emphasize their role as "information carriers" over any physical format. He argues that documents are tangible objects that act as evidence provision systems working within and underneath potentially biased systems of interpretation. Buckland maintains the need for an expanded conception of what a document is, calling for a definition that inherently takes into account that documents are not neutral carriers of data. Buckland ends by problematizing what constitutes data in the first place.

Sabina Leonelli's 2016 monograph *Data-Centric Biology* uses a science-studies perspective to first examine how biologists interact with data in a top-down manner, then further understand how specific data practices shape knowledge production in contemporary biology. Leonelli provides an in-depth philosophical study of how the life sciences have become data-centric. She examines how scientific knowledge is produced and validated through data collection, modeling, and circulation. She goes into depth about how biologists' interactions with data can introduce biases, as well as how technological tools can potentially worsen or mitigate such biases within the data as well as in the subsequent visualizations. Lionelli builds upon arguments made by Drucker and Hunsinger that "data" does not pre-exist, but instead is created through the work of humans (in this case, researchers), and as such should be analyzed with methods that include reflexive, humanistic perspectives. Leonelli argues for an approach that recognizes how data constructs knowledge while also retaining the importance of theory and qualitative analysis. Her work has significant implications for data epistemology across disciplines. Lionelli's book also provides empirical grounding for the application of something

like Hunsinger's theoretical framework of classification.

Paul Edwards' *A Vast Machine* (2013), traces the historical construction of computer-generated climate simulation models used to inform policy on anthropogenic global warming. Edwards argues that the scientific knowledge production techniques underpinning the climate change issue epitomize the modern reliance on computational models and data-centric practices. The book offers a genealogical analysis of the political and epistemological contexts behind these paradigms.

This deep dive into the practices and terminology conversations underlying climate change science offers insights applicable to research investigating how language choices shape data practices and vice versa. By highlighting the political dimensions behind authoritative reality-defining datasets or models, Edwards mounts a methodological argument about foregrounding context that resonates with goals to expose and potentially shift terminology conventions toward alternate information-type taxonomies. The essential linkages between model assumptions encoded in labels and consequent action pathways are particularly significant takeaways.

<u>Conceptualization, Interpretation, and Valuation</u>

Recent critical examinations of "data" have examined its nature - questioning conventional suppositions about its objectivity and meaning. These works highlight how defining and applying "data" as a term can shape knowledge production, as well as working to undo assumptions about the objectivity and neutrality of "data".

Johanna Drucker's writings probe how information is visually displayed and interpreted. In both "Graphesis" (2010) and "Humanities Approaches to Graphical Display" (2011), she contends data is not a raw material whose meaning is self-evident. She continues, explaining that

visual representations involve choices made by humans that can misrepresent ambiguity and uncertainty in data sets and therefore these representations are not objectively neutral or unbiased. She explains these choices by making a distinction between "data" and "capta", something further expanded upon in Hunsinger's paper. Drucker argues that visual representations should be understood as capta (taken) rather than data (given) since they are constructed interpretations rather than objective facts. She provides a critical analysis of how graphical forms can produce and constrain knowledge. This paper serves as a basis upon which to build out the idea of "capta" replacing "data".Drucker ends by championing the idea of "graphical excellence" - displays that inherently prioritize complexity, ambiguity, and subjectivity. Drucker further expands on the concepts laid on in "Graphesis" and its implications for the digital humanities in "Humanities Approaches to Graphical Display." She advocates for a humanistic approach to graphical display that focuses on interpretation, ambiguity, affect, and subjectivity rather than visualization models that mistakenly claim objectivity. This paper allows for further exploration of the practicalities of replacing "data" with "capta" and other, more descriptive, terms.

Published in 1979, *Laboratory Life: The Social Construction of Scientific Facts* presents an anthropological-style ethnographic study of the daily practices among a scientific laboratory community, in this case within the context of a biochemistry lab. Latour and Woolgar provide groundbreaking descriptive insight into the mundane culture underlying knowledge production rather than focusing solely on isolated endpoints.

The book frequently engages concepts like "inscription," "credibility," "cycles of credit," and the transformations of substances to instruments used in laboratory processes to eventually yield graphs, papers, and presentations. Through these processes, Latour and Woolgar emphasize

that terminologies are actively constructed by communities of practice.

Their detailed observations of laboratory activities uncover the myriad small judgment calls, "artificiality" tactics like metaphor or visual representation to demonstrate novelty, and rhetorical self-reinforcement necessary to establish credibility within the field to stabilize temporarilyy constructed facts. This behind-the-scenes view reveals the fragility and contingency behind truth claims taken for granted externally.

Latour and Woolgar's reveal of the social negotiations and terminological maneuvers involved even in supposedly detached lab settings bears strong applicability for literature arguing that data, evidence, and other information types gain meaning and force through conceptual labeling techniques. The observer lens modeling the collective construction of conventions for scientific fact-marking has clear relevance to this paper's aims in critically studying terminologies of information types today. Their study was also seminal in initiating critical data studies scholarship.

<u>Calls for New Frameworks</u>

Similarly to Drucker, Jeremy Hunsinger cautions against treating digital data as an objective truth. In his 2019 chapter "Big Capta?", he argues that even so-called "big data" must still be critically examined and interpreted. Hunsinger proposes a new linguistic framework to label and classify information, one piece of which is the aforementioned "capta" to denote information that has come into contact with humans before dissemination. Hunsinger also contends data has no inherent meaning and that there is inherently a human role in making meaning from data. He comments on the trend towards "big data" across fields and argues for the importance of critically examining how knowledge is constructed through data collection and analysis, rather than assuming "data" is an objective reflection of reality. Hunsinger uses a

theoretical framework to show how "capta" and other source/use-descriptive terms for

"information" can reveal meaning beyond quantitative approaches.

In a far-ranging conceptual analysis, Christine Borgman's *Big Data, Little Data, No Data*

(2017) confronts existing tendencies within technological solutionism that position "big data" as

a universally disruptive paradigm across scholarship and society. She surfaces critical tensions in

this monolithic framing of information, systematically delineating alternative constructs needed

to represent digital information diversity. Her central contention speaks directly to this paper's

aims - she advocates moving beyond broadly using "data" as an imprecise all-encompassing

term to categorize variegated, context-specific types of information.

Borgman grounds her argument by rigorously defining concepts like data, evidence,

records, documentation, and differentiating genres like log data, sensor data, personal data,

administrative data, etc. These conceptual bifurcations model the terminology diversification this

paper seeks to implement. By detailing data provenance theories and social roles, Borgman

builds her case for domain and community-specific information paradigms serving specialized

purposes, pushing back against assumptions that all data types possess commensurability or

interoperability.

A significant discussion contribution lies in delineating epistemological tensions between

traditional social science "small data" modeling approaches that privilege contextualization,

versus "big data" empiricism claims of letting large datasets speak for themselves without

hypotheses. Borgman explains that all information ecosystem elements require cultural and

historical interpretation, resisting technological solutionist tendencies. This creates space for

arguments to shift away from language emphasizing such inherent subjectivities.

Borgman constructs evolving taxonomies of information types, forms, purposes, and

significances centered on domains of practice like commerce, government, entertainment, and scholarship. These models inform proposals for new systemic vocabularies as alternatives to the umbrella "data" term. Borgman also addresses crucial distinctions around concepts like records and documentation that carry very different connotations than those found in popular data discourse.

While eschewing outright determinism about terminological shifts singularly transforming broader landscapes, Borgman concludes that resisting the generalization of data through alternative category-conscious constructs promises to enrich information policy debates, data literacy curricula, funding models, and design choices. This aligns with the current paper's aspirations. She calls for cross-disciplinary dialogue and recognition of tensions between qualitative paradigms and quantitative empiricism.

Borgman's lexicon models, conceptual precision, and illumination of competing epistemologies offer a strong foundation and parallel argument to this paper's proposed direction. Her elaboration of domain and community-specific genres resists assumptions of universal applicability and commensurability under singular data terminology. These moves open vital space for new vocabulary addressing the immense diversity of the evolving networked information ecosystem.

Cumulatively, these pieces look beneath the surface of data and documents to define them respectively as constructions of information interpreted through contextual human lenses, and the carriers of that information. There is a clear understanding that this information does not speak for itself objectively which is made explicit in multiple ways as the throughline of all these works. Although *these* perceptions about data are made clear, there is no universally applied solution for information classification, the definition of data, or how any of these changes might

shift our understanding or value of the information we receive.

Definitions

<u>Data</u>

Data in its true meaning is defined by Johanna Drucker in "Graphesis: Visual Knowledge Production and Representation" as follows: "Data are considered objective 'information'" (Drucker, 2010). Schroeder defines data as having three characteristics: "(1) it belongs to the object investigated; (2) it exists before analysis; and (3) it is the most divisible or atomized unit of analysis" (Schroeder, 2014). Hunsinger continues this extrapolation into data by saying that "data then is a part of something extant in the world, which can be analyzed, and has units" (Hunsinger, 2019). The inherency of data is its properties as purely objective information - being "given" instead of "taken". Ultimately, data is data because it cannot be anything else, it is the most granular piece of information possible, the singular elemental atom to the complex molecule that might be capta, acta, or any other category of information.

<u>Capta</u>

Drucker also defines capta in "Graphesis" as,"...information that is captured because it conforms to the rules and hypothesis set for the experiment" (Drucker, 2010). Through this definition, she provides us with a key element of what capta is at its essence. The human aspect of information, exemplified through the collection process, is what subjectifies data and makes it capta. In other words, capta is only capta because someone had reason to collect it within a certain set of parameters, whatever they may be. This definition further cements the objectivity of data, the subjectivity of capta, and the importance of the difference between them.

<u>Acta</u>

Acta is defined by Hunsinger in "Big Capta?" as "...acts or transactions including the

creation of an official record...acta implies something created as a record or a transaction. Acta then is information of record, where it represents the changes of the system" (Hunsinger, 2019) Hunsinger goes on to explain that currently, most data that records a change of state could, and should, be categorized as acta. He gives the example of databases that tend to exist for the sole purpose of recording changes of state and proposes to rename them "actabases"(Hunsinger, 2019). Acta is vitally important to the fabric of our society insofar as it records the abnormalities of a system which then allows us a basis for normative action and judgment in any applicable context.

Cognata

Hunsinger explains cognata's persistence as a term by saying that: "it…recognizes problems with data for anthropology and similar interpretive disciplines [and] acknowledges a mixing of cognition with information, much like there is with every analysis. Cognition is present in the creation, collection, and analysis of information" (Hunsinger, 2019) which opens up an entirely new spectrum of understanding for the categorization of information across the board.

Hunsinger continues by explaining that "by using cognata we would have to recognize much more substantively the work that forethought, afterthought, and thought, in general, brings about regarding observations and computations. Cognata is intertwined with…subjectivities in the past, present, and future and thus requires much more thought about its implications" (Hunsinger, 2019).

Communicata

Communicata is defined as "information arising from dialogue or discussions among people..." (Hunsinger, 2019). Most descriptions of communicata focus on examples like the

dialogue of a government panel, or conversations between people doing citizen science. Hunsinger further explains, "Recognizing the value of communicata as a dominant information type would help us overcome some of the traditional biases present in research, such as the fictum of the 'lone researchers' and associated topics...it is possible to see how it could aid significantly in helping the public understanding of science." (Hunsinger, 2019)

<u>Sumpta</u>

If data is defined as something that is "given " and capta is defined as information that is "taken" then sumpta is information that is given or taken as part of a bigger collection which becomes incidentally useful almost immediately (Hunsinger, 2019). Hunsinger explains this as follows: The usefulness of something is a matter of us taking and imbuing it with importance. "by making things useful, or by recognizing the usefulness of something, we transform it from something out there in the world called data, to something internal to us, which is called sumpta" (Hunsinger, 2019). The example that he gives is that of a random rock, versus your pet rock, the latter has been imbued with personal importance and has use, versus the former which simply exists (Hunsinger, 2019).

<u>Inventa</u>

Inventa is best defined by this quote from Schiller in his 1993 paper "Data, Datives, and Ablatives," "We shall have to accommodate also a variety of inventa, things stumbled or hit upon and found, though we may feel doubtful whether the mere fact that they are somehow there, yields us any guarantee that they can be put to essential uses." Inventa best describes information that is stumbled upon through "related/unrelated" methods of study. (Schiller, 1993) Hunsinger explains it as such, "Even though it has not been used extensively elsewhere, conceptually, it has possibilities to help us understand the processes of big data, which sometimes has inventa-like

findings"(Hunsinger, 2019). Inventa is never categorized as such when it is found at first but only becomes inventa once it becomes useful sometime after its initial discovery. The inherent nature of inventa is a stumbling upon of information which then leads to unforeseen usefulness.

Research Questions as Addressing Gaps in Literature

While the above works have established influential frames and philosophical groundings analyzing relationships between terminology, quantification, knowledge production, and governance, significant empirical and contemporary gaps persist. Examples include testing terminology effects experimentally, examining public divergence, tracing recent computational intensifications, confronting cross-disciplinary divides, and constructing alternative category schemes. Each gap offers vital avenues that the proposed research questions directly engage through mixed-methods inquiry.

<u>Lack of Experimental Testing</u>

Extensively cited milestone texts ranging from Desrosières' genealogy of statistics intertwined with state power, to Drucker's theorization of humanistic data interpretation, have richly contextualized the contingencies, subjectivities, and assumptions embedded within terminology choices past and present. However, few studies expressly implement experimental designs measuring the impacts of terminology itself as an independent variable upon other outcomes like understanding, valuation, or question formulation. While the likes of Borgman, Leonelli, boyd, Crawford, and others surface taxonomy tensions between computational and interpretivist paradigms through case studies, ethnographies, and critique, directly testing mechanisms of terminology causality could further enrich this foundation.

By taking an explicitly experimental approach toward evaluating terminology effects, this project's aims gauge both quantitative impacts on valuation, as well perceptions and attitudes via

qualitative studies, together assessing conceptions of data, capta, and proposed categories. Study designs span surveys, computational linguistics, creativity tasks, and wiki content contribution experiments that deliberately expose participants to terminology shifts. The combination of empirical observations, statistical analyses, and grounded qualitative themes linking terminology selections with impacts on knowledge production circles directly addresses existing scholarship gaps, begging further elucidation of terminology's causally constitutive role.

Public Discourse Divergence

Another gap lies in the analysis focused almost exclusively on academic debates and specialized disciplinary discourses around contested data constructs, overlooking terminology conceptions and literacy among the general public. Scholars continue to thoroughly unpack terminology tensions within the philosophy of science, critical data studies, science and technology studies, and similar domains while minimally inspecting wider public assumptions, attitudes, and codifications potentially diverging from said cutting-edge debates. This project aims to confront such imbalances by identifying patterns that hold significant implications for data literacy initiatives seeking to synchronize lexicons enabling democratic participation.

Reifying Forces of Computational Contexts

While the historical contingency of mathematical probability, accounting statistics, and other data precursors gained lucid excavation by canonical scholars like Hacking and Desrosieres, the intervening decades of exponential computational progress demand refreshed contemporary analysis deeply investigating terminology's coevolution amidst digital contexts. None of the sampled literature distinctly traces or closely examines the specific role cutting-edge advances like widespread sensors, AI algorithms, social media platforms, and other socio-technical shifts may play in embedding and amplifying terminology status quos. This

research plans exploratory archaeology scoping the magnitude of influence exerted by such recent technical accelerants through research questions probing why and how "data" terminology attained ubiquity across spheres, alongside how computational relationships may interplay with terminology choices about new information like automated predictions

By interweaving media discourse analysis and data infrastructural ethnographies with the emergence of research subfields, dual lenses elucidating both external terminology diffusion patterns and internal technical developments allow pinpointing key computational junctures. Machine learning techniques' promises to infer non-obvious insights from exponentially amassing quantified datasets likely reinforced conceptions that valid, valuable knowledge hinges upon continuous digitization into "data." Technological solutionist language permeating innovation rhetorics potentially further obscured nuances between data, capta, and documentation. Tracing and disentangling such complex chronological forces promises deeper clarity around terminological entrenchments.

<u>Cross-Disciplinary Terminology Divides</u>

Further opportunities lie in expanding beyond isolated disciplinary terminology investigations toward intentionally multi-epistemic examinations of cross-cutting data debates including dialogue between scientific paradigms and humanistic traditions. While scholars like Drucker, Borgman, and boyd thoughtfully engage some tensions between natural science and social science worldviews regarding information categories, few expressly synthesize perspectives simultaneously from information studies, quantitative computer science, AI ethics, and other domains that approach data production opaquely. Each conceptualizes terminological issues distinctly while rarely directly addressing divergence or opportunities for reconciliation.

This project purposefully breaches barriers via mixed methods spanning a selection of

fields coupled, with research questions eliciting commentary contrasting researchers and publics. Interviews, surveys, and embedded observations situate questions about terminology impacts within explicitly cross-disciplinary environments, refusing false binaries that narrowly define epistemic camps. Exploring terminologies as boundary objects illuminated differently across knowledge production communities promises meta-perspectival syntheses yielding new insight into entanglements between competing vocabularies and constructs that jointly shape modern information ecosystems.

<u>Constructing Alternative Information Lexicons</u>

Finally, a yawning gap demanding urgent scholarship lies in the practical task of constructing alternative vocabulary systems as taxonomies categorizing variegated information types contextually. Myriad existing works convincingly reveal "data" terminology limitations for precise communication about the emerging diversity of data-centric genres and formats amidst calls for enhanced literacy. Yet where critics thoughtfully dismantle existing language conventions, the proactive proposition of comprehensive replacements or supplemental lexicons remains lacking. This project responds through research questions designed to directly evaluate reception to formal category schemes for delineating forms of information based on situated roles and relationships beyond the constraints of reductionist "data" language.

The mixed methods provide integral scaffolding to organically foster alternative terminology. While results resist generalizable unilateral terminology shifts, the inquiry intends to richly populate starting conceptions for enhanced, precise language better representing information diversity. In centering the development of missing lexicons, this research addresses a profound gap by moving from deconstruction to creative regeneration.

In totality, the proposed original research questions, encompassing terminology impacts

on understanding, valuation, and question formulation directly build upon and integrate gaps in current literature. Through qualitative, quantitative and, experimental projects situated within and across these gaps, the resulting terminology inquiry promises multifaceted, contemporary insight enabling 21st-century data literacy improvements fitting information diversity. With rigor and creativity befitting the complex challenge, this undertaking transforms identified gaps into foundational areas for vital terminology negotiations toward knowledge equity.

Proposed Methodology

This project centers a critical paradigm oriented toward questioning information schema, challenging knowledge politics, and dismantling dominant discourses that may constrain terminological conventions marginalizing certain information forms and communities. These emancipatory aims seek to diagnose issues within existing data terminology frameworks that obfuscate and disempower through a lack of literacy, transparency, or diversity. The orientation aligns with critical data studies scholarship that applies reflexive, interpretivist lenses highlighting researchers' non-neutral role in methodology choices and knowledge production. With foundations in the influential scholarship of Foucault, Bourdieu, Haraway, and others, this critical framing informs selections of methods attuned to implicit hierarchies, biases, and rhetoric within technical systems. Quantitative computational techniques often positioned as objectively definitive get contextualized rather than reified through integrated qualitative questioning of underlying assumptions encoded in practices. No single category of methodology holds a monopoly on productive criticism or reformulation of terminological tensions.

At the same time, critique risks falling into inaction without coupling deconstruction efforts to applied, creative regeneration of potential solutions, however contingent or situationally embedded. Therefore, the methodology also embraces a participatory orientation

seeking to engage stakeholders across disciplines and sectors in collectively envisioning and prototyping possibilities for enhanced conceptual clarity. The integration of computational data gathering, humanistic questioning, and community-engaged design all provide complementary strengths suiting the complex challenge of evolving new shared vocabulary.

In service of bridging critical and participatory commitments, the methodological foundation of this project lies in mixed methods that systematically triangulate qualitative depth, quantitative breadth, computational analysis, participatory co-creation, media discourse views, and other angles illuminating multifaceted terminology questions. No single methodological orthodoxy intrinsically outweighs others' legitimacy per the paradigmatic positioning. Qualitative, quantitative, and experimental approaches get strategically combined for holistic assessment.

For instance, computational techniques like machine learning discover themes and clusters within corpora of texts while expert interviews probe specialists' nuanced reasoning. Quantitative multivariate statistical tests assess terminology perceptions alongside close discourse analysis of historic published debates on definitions. The integration intentionally interleaves empirical observations, logical inferences, reflective interpretations, and creative speculations without solely privileging a single knowledge production form.

In focusing extensively on terminology itself as the core subject of analysis, this project foregrounds often backgrounded functions of language infrastructure underpinning modern information ecosystems. Methods encompass not just direct investigation but performative, arts-based techniques that heighten awareness of taken-for-granted vocabulary grounds enabling certain worldviews. These methods aspire to render visible and open to reconfiguration embedded cognitive frames reinforced through terminology status quos.

The methodological scope spans academic, public government, non-profit, and industry spheres through multi-sector sampling and analysis attentive to potential divergence. Historical trajectories, contemporary cases, and speculative futures inform the temporal range of questioning centered on the decades of pervasive digital computation escalating datafication. Geographically, the project encompasses terminology conventions in Western advanced information economies as standard-setters with global influence while acknowledging situated specificity that resists outright universal generalizability.

In focusing extensively on the relationships between terminology and resulting explanatory powers, interpretive choices, or valorized information forms, the methodology directly accounts for positionality via my background, knowledge base, and intersectional identity as shaping inquiry framing and priority focus areas. Rather than seeking detached neutrality, critical reflexive examination of assumptions underlying choices of questions, sampling, comparisons, and vocabulary provides transparency regarding lenses applied. Citation politics also counter hegemonic erasures by foregrounding historically marginalized contributor perspectives integral to contemporary data debates.

No method inherently maintains immunity from the terminology paradigms under investigation - even apparently extractive computational analytics get contextualized rather than depoliticized given techniques' entanglements amidst commercial big data platforms and state surveillance capitalism critiqued by the project's critical paradigm. Ultimately the mixed methodology intends not positivist proofs but generative questioning across paradigms toward terminology negotiations better reflecting contemporary information diversity. The underlying power analysis seeks integration rather than domination of knowledge production modes.

Ultimately, this project's critical yet participatory methodology paradigm informs and

enables a mixed methods scope encompassing myriad approaches to terminology interrogations across sectors, time periods, and geographies. The research undertakes a reflexive examination of positional assumptions while directly investigating language politics and opportunities. Through qualitative depth, quantitative breadth, computational analysis, and participatory negotiation, the methodology framework befits the complex challenge of evolving shared vocabulary for more precise, equitable modern information ecosystems. The complex task of developing alternative terminology fitting contemporary contexts lies beyond any one field, yet currently lacks shared governance or infrastructure. More work remains ahead, but this project's methodology lays a foundation for continued conceptual evolution through collaborative terminology development, unraveling reductive assumptions hindering comprehensively ethical, accurate, modern information ecosystems.

References

Borgman, Christine L. *Big Data, Little Data, No Data: Scholarship in the Networked World*. MIT Press, 2017.

Buckland, Michael K. "What Is a Document?" *Journal of the Association for Information Science and Technology* (1997). https://people.ischool.berkeley.edu/~buckland/whatdoc.html.

Desrosières, Alain. *The Politics of Large Numbers: A History of Statistical Reasoning*. Harvard University Press, 1998.

Drucker, Johanna. "DHQ: Digital Humanities Quarterly: Humanities Approaches to Graphical Display." *Digital Humanities Quarterly* 5, no. 1 (2011). http://www.digitalhumanities.org/dhq/vol/5/1/000091/000091.html.

———. "Graphesis." *The Journal of the Initiative for Digital Humanities* 2, no. 1 (December 10, 2010). https://paj-ojs-tamu.tdl.org/paj/article/view/4.

———. "Non-Representational Approaches to Modeling Interpretation in a Graphical Environment." *Digital Scholarship in the Humanities* 33, no. 2 (July 7, 2017): 248–263. https://doi.org/10.1093/llc/fqx034.

Edwards, Paul N. *A Vast Machine: Computer Models, Climate Data, and the Politics of Global Warming*. MIT Press, 2013.

Hacking, Ian. *The Emergence of Probability: A Philosophical Study of Early Ideas about Probability, Induction and Statistical Inference*. Cambridge University Press, 2006.

Hunsinger, Jeremy. "Big Capta?" In *Springer eBooks*, 767–784, 2019. https://doi.org/10.1007/978-94-024-1555-1_16.

Kelty, Christopher. "Two Bits: The Cultural Significance of Free Software." *Choice Reviews*

Online 46, no. 04 (December 1, 2008): 46–2158. https://doi.org/10.5860/choice.46-2158.

Latour, Bruno, and Steve Woolgar. *Laboratory Life: The Social Construction of Scientific Facts*.

    SAGE Publications, Incorporated, 1979.

Leonelli, Sabina. *Data-Centric Biology: A Philosophical Study*, 2016.

    https://dx.doi.org/10.7208/chicago/9780226416502.001.0001.

Perry, D'Hana. "All Data Is Capta." *Data Visualization & Information Aesthetics*. Last modified

    December 6, 2018. https://dvia.samizdat.co/2018/all-data-is-capta-2/.

Schiller, F. C. S. "Data, Datives, and Ablatives." *The Journal of Philosophy* 30, no. 18 (August

    31, 1933): 488. https://doi.org/10.2307/2016787.

Schroeder, Ralph. "Big Data and the Brave New World of Social Media Research." *Big Data &*

    Society 1, no. 2 (July 1, 2014): 205395171456319.

    https://doi.org/10.1177/2053951714563194.