Sensing the Cloud: A Materialist Spatial Analysis of Data Centers and Critical Conceptualization

Tyler McCrea

A thesis
submitted in partial fulfillment of the
requirements for the degree of

Master of Arts

University of Washington
2021

Committee:
Bo Zhao
Sarah Elwood

Program Authorized to Offer Degree:
Geography

University of Washington

Abstract

Sensing the Cloud: A Materialist Spatial Analysis of Data Centers and Critical Conceptualization

Tyler McCrea

Chair of Supervisory Committee:

Bo Zhao

Department of Geography

With the ever increasing digitization of the global economy and everyday life, it is essential to understand the distribution and impacts of networked infrastructure, particularly Data Centers. While Data Centers as infrastructural objects have existed in some form or another since the advent of the internet, over the course of recent decades there has been a pronounced shift in the number of data centers being constructed, a marked increase in building size, and resource use, and dramatic changes in the types and purpose of data centers, particularly pronounced since the advent of Cloud Computing and the growth of the Public Cloud. This thesis provides an initial investigation of data centers in the United States and consists of two interrelated parts. I first conduct an exploratory spatial analysis to examine multiple aspects of the data centers regarding

their spatial distribution, position within the built environment, and land-use trends. Analysis revealed the highly clustered nature of data centers within metro areas associated with other physical infrastructures such as highways and land-use change agents. Second, with the results drawn from the spatial analysis, I conceptualize data centers as materially and spatially relevant nodes to comprehend the socio-technical, political, and economic implications of data center infrastructures. I root this infrastructure solely within the means of capitalist production and accumulation.

Table of Contents

Part A - Introduction, Literature Review & Research Questions

Part B - Materialist Spatial and Temporal Analysis of Data Centers in the US using GIS and Remote Sensing Methods

Part C - Space, Time, Extraction & Accumulation: Toward a Conceptualization of Data Centers as Multi-Scalar Nexus of Technocapitalist Relations

Part D - Conclusion

Appendices

List of Maps

List of Figures

Part A: Introduction, Literature Review, and Research Questions

"Darwin has interested us in the history of Nature's Technology, i.e., in the formation of the organs of plants and animals, which organs serve as instruments of production for sustaining life. Does not the [study] of the productive organs of [people] in society, of organs that are the material basis of every particular organisation of society, deserve equal attention?"

"Technology reveals the active relation of [hu]man[s] to nature, the process of production...and thereby also lays bare the process of the production of the social relations of [their] life."

(Marx, Capital Volume 1, Part 4, Chapter 15)[1]

1. Project Introduction

Here in Seattle, Washington, the sky is cloudy, a misty rain has been falling since early in the morning, 1,600 miles away, in Council Bluffs, Iowa clear skies, the temperature is a crisp 48 degrees. As I write this sentence in Seattle, the words I am typing are making the 1,600-mile journey from my computer to a server rack in Council Bluffs in around 29.5 milliseconds.[2] Using Google Docs as my word processor means that when my document saves, it is synced from my device to my Google Drive account, the data is hosted inside one of Google's 12 US data center locations. Intuitively one would assume that my data would sync to the nearest data center to my location (located in The Dalles, OR). Instead, it is sent from Seattle to Iowa, routing through 16 connection points as it makes its way from my device to the sprawling 407-megawatt data center Google operates in an industrial park just south of Council Bluffs. Inside the data center, my

---

[1] These quotes have been edited to remove the gendered language/pronouns present in the original text. I feel these changes do not alter the meaning or intent of these passages, and my decision to edit them comes from a sincere desire to utilize this text in an inclusive manner.

[2] I determined the location of where I presume my data is being stored by using the 'ping' command through the terminal on my computer to determine the IP address of the google docs URL, which gave me the IP of 74.125.76.189. Using the ipstack (https://ipstack.com/) geolocation API, I was able to determine that this IP address belongs to Google, and is assigned to a computer in Council Bluffs, IA.

words make their way in the form of bytes to one or several server racks, churning above an elevated metal platform within the 115,000 sq ft facility (Google, n.d.).

Data Centers are collections of servers that process and store data. They are a vital part of Information Communications Technology (ICT) infrastructure and work in conjunction with other infrastructures (i.e., fiber optic cables, cellular towers, software, and network protocols) to enable digital communication, data transmission, and storage (Pradhan et al., 2018). Data Centers range in building size from several square feet to more than 1 million square feet. The *International Data Corporation* (IDC, n.d.) defines five 'types' of data centers: "Server closet, Server room, Localized, Mid-tier, and Enterprise." Data center types are classified by overall size, the cooling mechanism employed, or energy consumption. (CBInsights, 2019; Shehabi, Masanet, Price, et al., 2011). Mid-tier facilities consume between 1-10 megawatts, enterprise data centers range between 10-250 megawatts, and emergent hyperscale data centers consume around 250+ megawatts (Klemick et al., 2017). Over a day, a server closest type data center can consume as much electricity as one household, while a hyperscale data center can use enough electricity to power 82,000 homes (EIA, 2015).

It is challenging to know the exact number of data centers that exist. There are approximately 8 million total data centers globally (Li, Wang, Luo & Pan, 2019), with most data centers classified as 'server closets' and 'server rooms.' The focus of this analysis is on Enterprise-scale, and new "Hyperscale"[3] data centers the former is estimated to be several thousand in number (Holt & Vonderau, 2015), while the latter ranges between 400-500 facilities (Steffens & Comerford, 2018, JLL, 2018), these are the data centers which compose 'the cloud.'

---

[3] Hyperscale data centers being those which have more than 5,000 servers, these are usually part of the 'public cloud'.

It has been predicted that the number of enterprise and Hyperscale data centers will grow significantly along with the demand for data processing and storage (CBRE, 2021). The growth in the number of data centers and the steady increase in the physical size and power consumption patterns has coincided with and enabled the adoption of digital technologies in a range of industries and has driven the process of the “digitalization of life” (Morsello, 2016).

Data centers relate to and act as fundamental components within ‘the cloud.’In practice, the cloud simply describes computer networks that are composed of many interlinked data centers. Users can purchase computing time from cloud providers to run applications, store data, or host files. This differs from colocation data centers where firms own their servers and rent space to locate that equipment. While technically, the cloud is always composed of data centers, not every data center is a part of the cloud. That being said, privately run data centers are becoming rarer and rarer, and most colocation data centers, even if they are not owned by a cloud provider (Amazon Web Services, Microsoft, etc.), are linked to cloud data centers and act as partial ‘cloud nodes.’ Following the trend in the research to use these terms interchangeably (Hogan, 2018; Pickren, 2014; 2018; Furlong, 2020; Amoore, 2016), I utilize ‘the cloud’ to describe the object of analysis in this paper, which is at once cloud computing data centers, and their linked colocation counterparts.

The seemingly all-encompassing nature of the trend of digitalization has resulted in an ecosystem of “airborne” data, wherein the lines between bodies, technology, and data blur. It has become hard to determine where data is captured, processed, and stored and for what purposes (Hogan, 2018). Data centers are the material space where the airborne data of wireless communication is brought down to earth and stored. The distinctly material locations comprise the imagined immaterial “cloud.” Data centers are sites of data collection, processing, and storage, the locations where ‘raw’ capital (i.e., space, data storage, data processing) is transformed, commodified, and valorized (Sadowski, 2019; Crain, 2016).

Data centers operate as a productive “organ” for the capitalist class (Marx, 1977), one of many that compose the wireless communication and computing networks that have come to undergird the modern economy. If (1) study of technology can reveal the processes of production and relation of humans to nature as Marx suggests, and (2) the study of the productive *organs* of society is essential to understanding the dynamic between technology and social relations, then it follows that contextualizing data centers and their role within capitalist production systems is critical to understanding shifting social relations.[4]

This thesis examines data centers’ spatial and material characteristics in three US metro areas (Northern Virginia, Chicago, IL & Phoenix, AZ) through spatial and temporal analytical methods using GIS and remote sensing technologies. Building from that analysis, I conceptualize data centers as multi-scalar nodes within networks of racialized technocapitalist relations. This study aims to build upon previous studies related to the social and economic impacts of data centers by demystifying the modern data center, making visible cloud infrastructures, and understanding how modern data centers operate as sites of extraction and accumulation by synthesizing and adapting multidisciplinary theories to understand better the unique spatiality and dynamic characteristics of this specific infrastructural “organ.”

---

[4] Originally quantifying and enumerating the myriad environmental/ ecological impacts, and varied emissions from data centers was a driving motivation for my research on this topic. While I do discuss this aspect of data center research in the literature review, and elsewhere, it is not the dominant thread of this project, and lies beyond the scope of this paper.

2. Literature Review & Areas of Contribution

What follows is a review of the current literature related to the study of data centers, the disciplinary theories, and the bodies of knowledge that shape this research. I note some opportunities for new contributions to these research threads and list the specific research questions that drive this project.

Existing socio-technical research has found that (1) technology and society interact in various ways[5] (Marx, 1977; Gideon, 1948; Latour 1999), (2) networks have a unique spatiality and materiality (Graham & Martin, 2001; Graham, 1998), and (3) that algorithmic violence is enabled and produced through communications infrastructure (Safranski, 2020; Rodgers & O'Neill, 2012).

It is also clear that data centers are increasing in size and scope (Shehabi et al., 2018; CISCO, 2020), though there remain lingering questions about where exactly data centers are being built and their spatial characteristics. There is an emerging need to study infrastructure's materiality and orient research within systemic processes and political realities. From these two positions, this project takes shape, which lies at the root of the formulation of this research.

I have divided this literature review into several categories, which layout the basis for this line of inquiry and contemporary research into data centers and other networked technologies. I argue for the necessity of this project. The categories for the literature review are as follows: *(1)* Science, Technology and Society (STS), *(2)* Materialist Geography, *(3)* Critical Data Studies, and *(4)* Critical Data Center Research.

---

[5] This is a bit of an oversimplification, and further attention will be paid to the differences between Feminsit, Marxist, and actor-network theories within STS in the literature review. Mainly focusing on how different scholars approach the ideas class oriented analysis, and agency/co-production arguments.

2.1 ***Science, Technology, and Society***

Research focused on the intersection of technology and society far predates the advent of digital scholarship (Gideon, 1948). Much of this scholarship belongs to the field of Science, Technology, and Society (STS). STS is an interdisciplinary field of study concerned with the material and social conditions under which scientific knowledge and technology are produced, and the implications of adopting said technology (Woolgar, 1991).

In many ways, the work of STS scholars shapes the ways we interact with, talk about, and understand technology today. This scholarship includes work on the nature of scientific revolutions and their impacts on society (Kuhn, 1962; Shapin 1996), the production of scientific knowledge (Knorr-Cetina, 1999; Gibbons, 1994), the construction of scientific "culture" (Latour, 1999; Collins & Evans, 2002; Pickering, 2010), a critique of modernity (Latour & Porter, 1993; Jassanof & Kim, 2015), and a current focus on digitization and its effects on society (Morsello, 2016; Reinsel, Gantz & Rydning, 2018). While a focus on technology and society is the overall throughline of STS research, numerous analytical approaches encompassed in this field.

One primary division in the field of STS lies in the distinction between technological determinists and scholars who argue that society shapes technology. Technological determinists view technology as a key governing force in society (Roe Smith & Marx, 1994); they posit that technological innovation follows an 'inevitable' course. That technological development determines social and cultural changes (Bimber, 1998). "Social shapers" (Salazar-Acosta & Holbrook, 2008), on the other hand, are those who view technology as being shaped by society. Several different approaches fall under this umbrella, including social constructivism (Pinch, 1996; Latour, 2003), Actor-Network Theory (Cutcliffe, 2000), along with Feminist STS

(Haraway, 1991; 1997; Harding, 2004; Eubanks, 2016), and Critical/Marxist positions (Klein & Kleinman, 2002; Hardt & Negri, 2004; Feenberg, 2014).[6]

Within the social shaping of technology approaches, there are also distinctions between scholars who subscribe to the Actor-Network Theory and those who are more aligned with Feminist and Marxist modes of analysis. However, these approaches do share many commonalities. The main difference between these theories is the role of agency and the influence of structural power dynamics. ANT scholars focus on the dynamic relationships between human and non-human actors. These networks of relationships are called assemblages (Muller, 2015), and ANT is interested in these assemblages' material and semiotic qualities (Law, 2008). Broadly ANT scholars reject distinctions such as subject-object dualism and treat human, non-human, and technological actors as equal participants in their mutually constitutive networks.

Marxists and Feminist STS scholars reject the aspects of ANT, which disregard context and structural power dynamics (Brenner et al., 2011; Sayes, 2017). Breaking with ANT in this way, Feminist STS scholars have focused their analysis on the development and adoption, and impacts of technology within patriarchical and racist systems (Shaw, 2019, and Marxist scholars centering the primacy of class struggle, captialist productive forces and the linked processes of matrial and social production (Goto, 2013). In response to these varying approaches and the drifting of STS scholarship away from Feminist and Marxist approaches, Arboleda calls for scholars to "Occupy STS!" stating that STS scholarship contemplates technology without

---

[6] Some scholars lump all of the approaches I called "social shapers" as belonging to social constructivism, I use Acosta and Holbrook's term of social shapers since it allows for more specificity when discussing the different approaches. Another important concept/theory which is also associated with social constructivism is co-production, which posits that scientific ideas and technologies evolve together through representations, identities, discourses, and institutions (Jasanoff, 2004).

problematizing its context. Insisting that study of technology is empty if it does not consider that these technologies are interwoven with relations of class and production (Arboleda, 2017).

This work is deeply informed by STS scholarship, and I am mainly motivated by Arboleda's call. Through a critical mapping and analysis of data centers, I will situate this technology within the material and social conditions of the contemporary US, a neoliberal capitalist state with a privately planned and owned telecommunications network, and examine the "politics" (Winner, 1980) and configurations of data centers as infrastructural objects. By orienting the methods of this project to have a spatial framework and considering the political and theoretical milieu from which this technology emerges, this project enriches the work of STS scholars and geographers. Working on describing the distribution of a given technology, situate the data centers within the capitalist mode of production and social relations.

2.2 *Materialist Geographies*

There have been several so-called 'turns' within Geography, such as the cultural turn in human geography (Barnett, 1998) or the visual turn (Thornes, 2004). Of direct relevance to this research are the recent "digital turn" in human geography evidenced by the growth in research on geographies produced *of*, *by,* and *through* the digital (Ash et al., 2016) and the "materialist" turn in human geography, exemplified by various overlapping engagements with materialism and materiality in geographic scholarship (Kirsch, 2013; Whatmore, 2006). When used in this way, Materiality and Materialism are two separate but complementary concepts, both encompassed by Materialist Geographies.

Materiality is so broadly used and employed in many different ways that it almost defies a simple definition. In particular, geographers have struggled with making sense of multi-faceted and overlapping materialities (Anderson & Wylie, 2009). In the broadest sense, materiality is the quality of being 'matter' (Froman, 2021), or put more simply, "the stuff of the world, both animate and inanimate" (Rogers et al., 2013). This still seems insufficient, and I refer to the distinction made by Kirsch between various material approaches, namely "gritty" materiality, which focuses on waste and the materials which compose, and "transition" materiality, which engages with the transformation and processes of materials (Kirsch, 2013). I am not seeking to put forth a definitive definition of materiality in any sense. Still, in the case of this analysis, I will use it to mean the materials both isolated and connected that are composed in or through the data center, from the raw materials utilized in building construction, or server creation, along with the land and soil on which the data center is built.

While materiality is more of a state of being or object, Materialism is a framework that is deployed in several different ways within geography, the two most prominent approaches are 'new materialism' and 'historical-geographical materialism' (Swyngedouw, 2003). The distinctions between the two are relatively reminiscent of the divide in STS, with new materialist approaches oriented towards exploring the social significance of materiality and incorporating Actor-Network theory, using assemblages, interconnected, dynamic networks of relations, as a way to study materiality (Forman, 2020; Tolia-Kelly, 2012). Historical-geographical materialism is an explanatory approach that considers present material conditions and draws upon concepts such as historical and dialectical materialism to understand the material conditions as fundamentally dynamic with the mode of production (Kirsch, 2020). Geographical Materialism, as employed by Soja, refers to Marxist spatial analysis, which considers how the production of

Space mirrors other social forms (the spatial problematic) and examines space and materiality through the process of reproduction (Soja & Hadjimichalis, 1979; 1985).

2.3 *Critical Data Studies*

Continuing the trend, another interdisciplinary field, albeit with a shorter history than STS, which informs and shapes this work is Critical Data Studies. This field of research emerged from conversations and general questions posed by two Microsoft engineers regarding the nature of big data and its implications for society (Boyd & Crawford, 2011). From these questions, emerged Critical Data Studies (CDS), the term was coined by Dalton & Thatcher (2014) in a piece which sought to lay out the critical framing questions for interrogating the role of new "Big Data Regimes," framing these regimes centers analysis on understanding how social relations are shaped by data collection and extraction (Zhang et al., 2020). Since that time, CDS scholars have explored the politics and power embodied in Big Data (Iliadis & Russo, 2016, Thatcher et al., 2018), the spatial implications of algorithm analytics (Dalton et al., 2016, Kwan, 2016), and examined the new futures opened by emerging geographical perspectives of radical politics and critical data inquiry (Burns et al., 2015), and the material nature of data processing, by exploring the energy consumption patterns of Bitcoin mining and the ecological impacts stemming from cryptocurrency mining (Lally et al., 2019).

Critical Data Studies emerged directly from considerations around collecting, processing and, aggregating data and the socio-political and economic implications of Big Data regimes. From that framing, CDS scholars have tended to orient their focus on software and data processing, and the creation of "code space" (Kitchin, 2014) rather than physical infrastructure. This is a trend across most strains of digital geography, where scholars have focused on the virtual, and theoretical rather than material or infrastructural considerations (Kinsley, 2014). The key foundations from CDS that most inform this research focus on data collection and processing as a means of political power and the centering of the idea of ‘Big Data” as a new computational regime. By focusing on the infrastructure of data centers through a spatial and material lens, I will explore how the generation of big data has coincided with and been enabled through the construction of large, linked data centers.

2.4 *Critical Data Center Research*

While digital infrastructures are ostensibly a component of Critical Data Studies and Digital Geographies, there has been, until recently, little research focused explicitly on investigating data centers through a critical lens. Much of the existing literature regarding data centers comes from the disciplines of computer science and engineering and focuses on improving energy consumption efficiencies (Andrae & Corcoran, 2013, Andrae & Elder, 2016), data center demand networking (Shehabi et al., 2018, Mann et al., 2011), and data center engineering considerations (Bahari et al., 2016). These pieces share a common focus as the critical research, in that they investigate data centers, but missing from this scholarship is a consideration for the systemic socio-technical and material qualities of data centers. The remainder of this section will focus on a review of the critical research on data centers, and synthesize the theoretical foundations of this line of research inquiry.

Tung-Hui Hu’s (2015) book *A Prehistory of the Cloud* is perhaps the starting point for the critical turn in data center research. In the book, Hu unveils how the infrastructure which constitutes “the cloud” is built upon past infrastructures, often non-digital infrastructure. He tracks how fiber optic cable networks have grown from early telegraph and telephone networks, which were usually constructed along the early railroads of the US. Another primary focus of this work is the connection between modern data centers and military infrastructure; he discusses the trend of data centers being built within former military bunkers and considers how the legacies of those past uses bleed into and affect the current use of these spaces as data centers. Hu contends that to fully understand digital infrastructures and their impacts, “we must begin with space, power, and the combination we call history” (Hu, 2015).

Louise Amoore’s *Cloud Geographies* (2016) is another excellent example of critical data center research that situates the relationship between cloud infrastructures and geopolitics. In the article, she notes the importance of mapping and knowing where the cloud is located and states that “Geography matters in the cloud” (Amoore, pp. 7). She also cautions that simply locating the cloud doesn’t reveal the intricacies of the connections between infrastructures and politics and must pursue an analytic which appreciates and incorporates context and political realities when studying cloud infrastructures.

More recently, there was an edited collection of truly paradigm-shifting scholarship related to data centers in the journal *Culture Machine* called “The Nature of Data Centers” (Hogan & Vonderau, 2019). This volume was co-edited by Mél Hogan and Asta Vonderau, two long-time critical scholars of data centers. This collection represents the growing importance of and focus on data centers as infrastructural objects as a focal point for critical research. Alix Jonson’s “Emplacing Data Within Imperial Histories” (Johnson, 2019) considers how Iceland has marked itself as a destination for data centers, advertising abundant, clean geothermal energy, but roots this development within a history of Iceland’s military-industrial complex and position at the periphery of global politics. From the same collection, Julia Velkova (2019) discusses the impermanence of data centers and the lack of what Shannon Mattern (2018) terms a "culture of maintenance." Velkova notes that most equipment within data centers becomes obsolete after 5-10 years (or sooner) and examines the waste produced due to the quick phaseout timeline of data center servers.

Another focus of the collection in Culture Machine is on the coproduction of data centers and nature, or wilderness. Levenda and Mahmoudi (2019) explored the connection between data centers, hydropower, and urbanization in the Pacific Northwest. They discuss how nature is constructed as both a resource and a greenwashing strategy for data centers and track the uneven spatial impacts of their construction. Additionally, A.R.E. Taylor discusses how data centers are idealized by tech firms as "Peopleless technological wilderness" (Taylor, 2019) and explores how "the infrastructure fiction of the depopulated data center intersects with fantasies and futures of technological progress, nonhuman security, automation and data objectivity." (Taylor, pp, 12, 2019). These pieces all draw from various theoretical foundations, including political ecology, ideas around sovereignty, and surveillance.

Graham Pickren's *Global Assemblage of Digital Flows* (2018) is another essential work on data centers, and the spatial problematic posed in the piece directly informs and shapes my research questions. In this piece, Pickren explores how the uneven distribution of computing resources often mirrors existing patterns of uneven development and also identifies key gaps in existing research around data centers, their spatial orientation, and the socio-political and environmental impacts of their operation:

> *Mapping out where data centers are, where and how they draw natural resources, and the economic, social, and environmental impacts on communities remains a crucial task*

> that has important implications in terms of understanding the relationship between computing and socio-natural change. (pp.237, 2018)

Pickren identifies the crucial role that data centers play within wider flows of capital and resource extraction and notes that we cannot adequately understand or assess the dynamics of those flows without a spatial grounding.

More recently, Kathryn Furlong published an article *Geographies of Infrastructure II: Concrete, Cloud and Layered (in)Visibilities* (2020) which terms the data center and associated infrastructure as “cloudfrastructures” and sets out to examine the existing research and center both the material and “more than material” impacts of data center operations. In her piece Furlong discusses the various component parts of infrastructure which define the cloud, and traces their physical and material costs while also examining how this infrastructure is purposefully mystified and rendered invisible; she calls this “Layered (in)visibilities” and notes that this is purposeful on the part of cloud providers. She concludes that more attention needs to be paid to not only data center materiality but the broader historical context of computational development, “paying attention to cloudfrastructures calls for new thinking on the axiom of (in)visibility in infrastructure studies” (Furlong, 2020).

From the review of the current literature on critical data center research, several areas for intervention emerge. The development of spatial knowledge around where data centers are located, and where they draw their resources from remains an important task. Additionally, attending to the dynamic between infrastructure and the socio-cultural and political is needed within research as the role of data centers within capitalist production systems is often obscured and masked by layered-invisibilities to serve a particular purpose. Importantly, the previous scholarship on this subject suggests that solely focusing on locating the data center is a mistake, we need to consider the infrastructure, but also to center the forces of capital and production which ultimately determine where a data center is built, who the intended users are, and for what purpose this technology is deployed.

2.5 *Areas for Contributions / Interventions*

Taken as a whole, the review above charts the intellectual and methodological foundations of this work from the broad ideas of Science and Technology Studies to the narrow thread of critical data center scholarship. This research is grounded in a rejection of technological determinism, centering instead critical and materialist modes of analysis to explore data centers through a spatial problematic and to frame their position within the capitalist mode of production. The interventions being made with this work are severalfold. First, most critical data center research has focused on individual data centers, usually highly unique examples in their design or operation. Here I will instead focus on data centers as a whole, focusing on the where and what of the "average" data center, to represent and visualize the current distribution and nature of these data centers as they're built right now. Taking up the call to restore the critical and radical STS research traditions I will as well contribute to materialist analysis by examining the multi-scalar nature of data center operations, as spaces of data capture and processing through the synthesis of existing social and economic theories.

3. Research Questions

To guide and develop this work, I have posed several research questions below and divided them into their respective chapters of this thesis. These questions guide the direction of the research methods and attend to the gaps and areas for intervention that were identified above in the literature review.

- ***(Part B)*** **Where and in what form/s can data centers be found?**
- ***(Part B)*** **How can spatial analysis be used to unmask the materiality of data centers, and what limitations exist to these approaches?**
- ***(Part C)*** **What theories or conceptual foundations can inform our understanding of data center materiality, as well as social, political, and economic entanglements at multiple scales?**

To attend to these questions, I begin with a spatial analysis of data centers at multiple scales and examine their spatial properties through spatial patterns and distribution, place-based analysis, and temporal land-use change. This analysis situates data centers through their spatial location, context, and material impacts and provides opportunities for reflection regarding the application of spatial analytical methods to study data centers. Informed by the results of the spatial analysis, I suggest three theoretical approaches for conceiving data center materiality at multiple scales, and explore how these theories can be used to unravel the dynamic relationship between data centers and their socio-spatial impacts, and material entanglements.

Part B: Materialist Spatial, Place-based and Temporal Analysis of data centers in the US using GIS and Remote Sensing Methods

Part B Abstract:

Exploring data centers through a spatial analysis remains an under-examined aspect of materialist geography. In the following sections, I present two approaches for such research. First, analyzing data center location data using GIS software to detect clusters and study the spatial distribution of data centers across three study areas. Second, by "Sensing" this infrastructure, detecting and studying the associated land-use patterns and changes of data centers using remote sensing techniques. Taken as a whole, this chapter examines the often muddled process of acquiring information about data centers, constructing a suitable dataset, what trends and patterns we can see in the construction and deployment of data centers across three regions, and in-depth exploration of temporal and spatial analysis of digital infrastructure using remote sensing.

The data center remains among the least studied areas of digital culture, with cloud computing producing a layer of abstraction that masks the physical infrastructure of data storage. (Tung-Hui Hu, 2015)

4. “The Blank Spaces” on the Data Center Map

Mapping, and more specifically locating data centers has been a challenge for geographers and others interested in the spatial contingencies of the cloud and network infrastructures (Burrington, 2016; Johnson & Hogan, 2017). The mapping gap identified by Pickren (2018) reveals that a lack of fixed spatial knowledge about the location of data centers precludes a cohesive assessment of the materiality of this infrastructure. A spatial understanding of data centers is important both conceptually and practically; it is foundational to materialist analysis and reveals the dynamics and processes between capitalist production and the (re)/production of space (Soja, 1985). Additionally, a spatial foundation is important to any sort of political organizing or praxis-based interventions into these spaces and the operation of the cloud as a space of production, surveillance, and accumulation. Without a spatial grounding and understanding of where a technology or infrastructure is built, where its users are located, and its material composition, we cannot properly understand the dynamics of economic or relational flows.

While the mapping gap identified by Pickren was the main motivation for the spatial portion of this research project, I am also drawn to the project of mapping data centers due to the previous work on this subject by artist and writer Ingrid Burrington. Burrington has published several articles and other works on the topic of data center and network mapping, which focus on tracking down data center locations using municipal documents, web crawlers, and news reports to map data centers and examine their associated material, political and economic dimensions

(Burrington, 2016; 2021). I am particularly influenced by the approach Burrington took for the art project “Reconnaissance” (2016) which used commercial satellite imagery to observe large infrastructure sites including military bases and cloud data centers. The goal of Burrington’s project was primarily around visualization and observability of data centers, rather than spatial analysis, but I am drawn to the idea of maps and aerial imagery as art (DOI, 2018), grounding visualization and exposure.

4.1 Approaches & Objectives

My intention in the following sections is foremost to explore the spatial aspects of data centers as a specific infrastructural object, but with key caveats. Firstly, I want to avoid the issue of spatial fetishism (Soja, 1985, Amoore, 2016) by stating plainly that in my observing and studying data centers through their locations, distribution, and building type, I do not believe that these data centers by their nature or operation are shaping the communities around them, or that data centers wield some unique power. Rather, data centers are ***where*** they are and ***what*** they are because of their position within the system of global capitalist production. I am observing and naming these places and their locations not in the hopes of understanding their power per se, but to better understand how power operates through these locations at various scales and to test spatial analytic methods in conducting materialist research around data centers.

In the following sections, I will attend to “spatializing” the data center through various spatial analysis and mapping techniques, working to locate and map data centers in order to examine the material nature of their construction and operation and to explore the spatial characteristics of this infrastructure. To do this I will first collect a dataset of data center locations in order to visualize the locations and to conduct an investigatory geospatial analysis using GIS software to identify clusters and explore the distribution of data centers across three study areas. Using the data center clusters I will examine the buildings themselves by exploring their "placial" characteristics. I also examine the community characteristics of the census tracts where data centers are located. Finally, I conduct a remote sensing-based temporal land-use change analysis, *sensing* the materiality of the cloud using satellite imagery, and visualizing data centers not as individual objects, but as a distinct land-use class.

What this project does differently from other mapping exercises is that it works at a larger scale than previous examples, and compares spatial patterns across study areas. Additionally, this work seeks to move beyond visualization for the sake of observability by integrating quantitative analytical methods along with mapping and visualization. Lastly, this work tests several methods and provides a starting point for further research and mapping activities. Below I will review the spatial and temporal analysis methods, findings, and synthesize why these approaches are important in filling the research gaps identified in the literature review.

5. Data Collection, Study Design, and Methods

5.1 *Spatial Analysis* <u>*or*</u> *Locating the Cloud*

5.1.1 Data Collection

Information related to data center locations, facility size, power consumption, and projected annual emissions is not easy to find (Kleyman, 2015, Pickren, 2017). Datasets are not readymade, and the data researchers are interested in collecting is dispersed across a number of industry adjacent websites, corporate reports, and technology forecasts. Tu, Hogan, Burrington and Furlong have all discussed the relative lack of information in the late 2000s-2010s when data center operations were largely kept secretive, and describe a corporate "opening" of data centers that has occurred recently (Hogan, 2018; Taylor, 2018). Despite the current relative availability of this data now compared to say 2015, this 'opening up' of the data center is more of a coordinated distribution of curated photos, and reports which are meant to present a sheen of both techno-futurist sensibilities and a "green smoke screen" to cloak or hide emissions-related environmental costs (Naschert & Tomaszczyk, 2020; Furlong, 2020).

Before devising specific study areas or methods, I set about looking for a dataset that contained information regarding data center locations, looking specifically for addresses or coordinate information so that the locations could be geocoded and mapped. Without a workable dataset of locations, I felt that this analysis would be stymied before I even started. I encountered a web of overlapping and incomplete datasets, what Furlong describes as layered (in)visibilities. Looking for a comprehensive dataset that included location as well as building size and energy consumption information proved fruitless. I decided to make do with the information that was available to me, and after reviewing several websites which were used to market data centers to prospective clients[7] I settled on what appeared to be the most robust (datacenters.com), and set about collecting information from the site to be compiled into my dataset.

I collected a dataset of 1,496 US data centers, which included the data center name, service provider (AWS, Google, Digital Realty, etc.) as well as the address, which was geocoded to retrieve the data center coordinates. This dataset was collected using a python web crawler (Thelwall & Stuart, 2006; Zhao, 2017) that crawled [8] the following website directory of data

---

[7] https://www.datacentermap.com/ | https://baxtel.com/map (two examples)

[8] "Crawling" here just means that each page of a given website is visited, and specific portions of the webpage are collected into a separate dataset.

center locations: https://www.datacenters.com/locations. Taking into consideration the number of queries needed and the amount of data to be collected, the crawler was made to be as unobtrusive as possible to collect the site information. I have stored this web crawler script in an open access Github repository to ensure access to others (McCrea, 2020).

While this initial dataset was a great starting point, there were issues from the start with data accuracy. The nature of colocation data centers[9] is such that sometimes there are multiple host providers within the same data center building which can lead to problems with duplicate or incorrect locations. Additionally, there was no way to easily collect the power consumption or building size data without significant formatting and data validation errors. Despite the high number of data centers returned by this crawler, this still was not all of the data centers in the US. The validation and cleaning process required cross checking locations and data center names with other datasets including (peeringdb, n.d. & udger, n.d.). This process caused me to develop a bit of an obsession with ensuring I was capturing every single data center, and trying to include every piece of data or statistic, which became reductionist in a sense. I fell into the trap of spatial/data fetishism fixating on miniscule data points, and locating every last center, which not only hindered my progress, but slowed the development of further methods.[10]

Taking stock of the dataset after much cleaning and processing, I realized that the scope of the analysis had to be simplified, and so I decided to move forward with the data that I was

---

[9] Colocation centers allow for multiple data center hosts to operate within the same building, and unlike cloud computing, every operator usually owns their equipment within the data center, so care must be taken to uncover who the overall operator for the center is, and to ensure that you are not double-counting locations where only a handful of servers is classified as a “data center”.

[10] Initially, I had grand plans to conduct a totalizing spatial analysis/review; mapping every data center in the US or even the world in an attempt to quantify the total emissions for all data centers, along with other environmental considerations. After coming to terms with my grandiosity, and taking an assessment of the time constraints of this research, I set about to conduct a smaller mapping project.

able to collect, incomplete though it may be. I shifted focus, using the location information that I was able to verify, and setting aside the energy and water consumption-related data that I had hoped to work with. The final dataset is over 1,200 verified US data center locations, which were geocoded using the geocodio.com service, and validated through random audit (out of 150 randomly picked addresses 89% were correctly located, with most errors only off by >50 meters). I also collected several other datasets to be used in the spatial analysis of these location points. These include Census Tract Shapefiles from the US Census Tigris geospatial dataset (US Census Bureau, 2018), a dataset of Amazon fulfillment center locations (MWPVL International, 2020), and a readymade dataset from the US Centers for Disease Control assessing Social Vulnerability at the census tract level (CDC, 2021). In the next sections, I will review the overall design for this study/analysis and the methods that will be employed to work with the datasets described above.

5.1.2 Study Design & Methods

Similar to Burrington's mapping practice, other scholars, artists and digitally minded researchers who have approached mapping and visualizing data centers have focused their analysis on one or two unique examples (Varnelis, 2014), oriented their studies to focus on one type of data center (i.e. Hyperscale) (Gihasi & Barca, 2014), or one provider's data centers (Data Farms, 2018; Mayer, 2019). While these are compelling studies, the narrow scope of said research tends to focus on examples that are outliers rather than the norm of most data centers, the vast majority of which have no special architectural features, and are meant to be subsumed into the background rather than stand out as spectacle. In this research, I will not limit the analysis to a handful of data centers or focus solely on AWS data centers, but rather on mapping all of the medium and Enterprise size data centers in a given study area, regardless of ownership. In developing my study and methods, I will not be looking to find the biggest or most energy dominant data center, or the most architecturally notable examples, rather I want to discern the ***where*** and ***what*** of the "average" data center to represent and visualize the current distribution and orientation of data centers within the built environment.

Study Areas

Despite initial intentions to conduct a country-level analysis of the entire United States, I refined the scope and defined three study areas to focus the analysis. These study areas were developed in conjunction with existing literature, largely from industry reports, about data center real estate patterns. I first selected a study area widely known to be saturated with data centers, an area of Northern Virginia known colloquially as 'Data Center Alley' which has been a large data center market since the early 2000s (St.Germain, 2019). This is an area I knew I would have an abundance of points for analysis, and so for the second and third study areas, I targeted a 'secondary' market (Chicago Metro) and 'developing' data center market (Phoenix Metro) to compare to the largest, and oldest (Northern Virginia). These study areas vary by scale in terms of land area, and population but each contains one or more counties, which will be divided into census tracts. The Northern Virginia Study area covers the counties of Loudoun, Fairfax, and Prince William; Chicago metro study area comprises Cook and DuPage counties; and the metro Phoenix area contains Maricopa county. These study areas cover a range of different geographical areas, and building and population densities affording an opportunity to see where patterns of development are similar or different, and in which ways.

Approach

There are three core components or elements of the spatial analysis which I conduct across the three study areas, and these are: *spatial distribution*, *place-based (placial analysis)*, and *census tract analysis*. To determine and explore the distribution, begin by mapping the data centers in each of the study areas, visually inspecting the distribution in relation to other pieces of infrastructure such as highways, and running a density-based cluster analysis on the data to determine if the data centers are clustered or dispersed. Building from the clustering analysis, I then look at the building types and forms of data centers by selecting one data center from every cluster and classifying the data centers into several building types based on their physical characteristics. Lastly, using the clusters I will subset the data to the census tracts to determine which tracts within the study areas are 'data center tracts' and analyze these tracts in relation to other infrastructure, in this case, Amazon fulfillment centers, as well as compare the Social Vulnerability of data center tracts to the overall study

Software, Tools & Methods

To conduct this study I relied mainly on ArcGIS Pro (v.2.7.0) to create static maps, and to conduct a clustering analysis. In the course of data collection, processing, and analysis I also utilized several other software tools including geocodio (geocodio.io, n.d.) to geolocate data center points, RStudio to view and perform preliminary data analysis using the ggplot and dplyr libraries, and QGIS to create test maps and preliminary visualizations. Ultimately ArcGIS Pro was selected for this analysis because of its wider suite of analysis tools, particularly density-based clustering methods, and speed of processing and map production.

The most computationally intense, and statistically significant mode of this analysis was the clustering analysis which was run for each study area's data center locations. In order to assess the distribution of the data center locations, to determine if there was clustering present, or if the locations were dispersed without discernible clustering. In order to determine the degree or presence of clustering, I had to decide between several different clustering approaches for this data. The two most appropriate approaches would be either a centroid-based clustering method, like K-means clustering (Likas, et. al, 2003), or density-based clustering such as DBSCAN or OPTICS. While each has its advantages, I chose the density-based approach for several reasons, first density-based clustering is able to detect noise, so while K-Means clustering assigns every point to a cluster, density-based approaches can detect clustering and noise, so not every point is assigned to a cluster (Ester et al, 1996). Additionally, density-based approaches can detect clusters of varying shapes and uniformity (Campello et al., 2013), so it is useful for identifying clusters that may vary in appearance such as with this data.

Expanding on the previous paragraph, after determining that a density-based clustering method was most appropriate, I had two options to choose from, each varying slightly. Defined distance (DBSCAN) and Multi-scale (OPTICS) density-based clustering are the two most prominent options, while DBSCAN uses a specified minimum distance between clusters and noise (Birant & Kut, 2007), OPTICS clustering measures the distance between each point from each neighboring point and creates a reachability plot to discern clusters from noise (Agrawal, et. al, 2016). The OPTICS method seemed the most applicable for this analysis as it did not require a defined search distance, and also allowed for both a minimum features per cluster parameter and a cluster sensitivity parameter which could be kept consistent across the study areas.

Running the OPTICS clustering method is fairly simple, for each study area the clustering algorithm was run with a minimum of 3 features per cluster, and a cluster sensitivity of '50'. Cluster sensitivity determines whether the algorithm is aiming to detect very few clusters (sensitivity closer to '0'), or many clusters (sensitivity closer to '100'). The algorithm outputs a reachability plot, which represents the distribution of points, plotted in accordance to their reachability distance, or the distance between a given feature and all other features within the search distance threshold (Agrawal, et. al, 2016). From this plot, clusters can be determined based on the shape of their distribution, with deeper valleys representing denser clustering of points. The reachability plots for this analysis will be discussed in more detail in the next section.

In addition to testing and visualizing the distribution of the data centers across the study areas, I will also attend to the other core components identified above: *place-based,* and *census tract level analysis*. To explore building form, I selected one data center from each cluster for a placial analysis. Using Google StreetView I went to the location of the data center and took a screenshot of the Street View image. Based on the images I classify the building types into three groups which are consistent across all study areas and explore the characteristics of these types. Lastly, I aggregated the data center location to their underlying census tracts to determine the number of data centers per cluster, and to identify which of the tracts could be termed 'data center tracts'. Using these tracts I conducted a simple distance-based analysis between the selected tracts and amazon fulfillment centers to determine overlap between these infrastructures, and also explored the social vulnerability of the data center census tracts, to the overall study area vulnerability, comparing simple summary statistics.

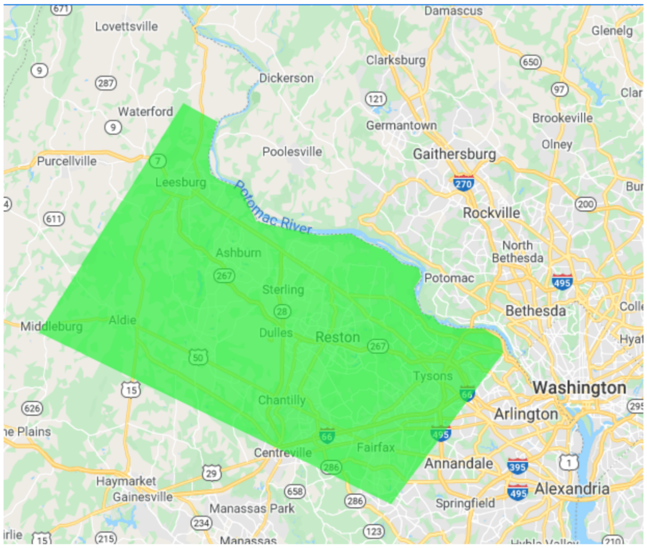
It is worth noting that this analysis is primarily exploratory in nature, rather than confirmatory. Focused on developing and testing multiple methods for spatial analysis pertaining to data centers, with an eye towards materiality, spatial distribution, and social dynamics. While I would have liked to present a much more comprehensive and wide-ranging survey of data center ecologies and environmental impacts, attending more specifically to scrutinizing all aspects of the material nature of these spaces, this analysis is one attempt to 'do the work' of materialist geographies, limited in scope that foray may be.

5.2 *Remote Sensing Analysis <u>or</u> Sensing the Cloud*

While the previous section was an exploration of how to work with data center location data for spatial analysis and constructing a usable dataset from diverse sources. This section considers how to contend with and overcome the "layered invisibilites" of the cloud (Furlong, 2020). For instance, Google and Amazon (AWS) will release data that shows improved efficiencies and carbon reduction, but data is released in bits and pieces, or abstracted as to belie its true nature, in other words pure "greenwashing" (Brodie, 2020), or what Furlong calls a "Green smoke screen" (Furlong, 2020). This can make it difficult or even impossible to fully understand the socio-environmental impacts of data center operations if such an understanding is possible in the first place.

While a robust analysis of the total environmental impact of cloud computing is beyond the scope of this or any singular study, in this section I lay out a methodology for 'sensing the cloud'. Whereby I conduct a temporal land use land cover change analysis for a given study area, with the aim of detecting and locating data centers, and revealing the scope of their land use, and associated land-use change over time for this infrastructure and its related infrastructures such as warehouses. This approach allows us to move beyond human vision and senses for seeing and hearing/interacting with the cloud, to viewing this infrastructure from beyond the clouds so to speak. We needn't rely on politicized, and cherry-picked industry data but can observe the material nature of the cloud and data center through other means. Below, I will review the data processing methods, analysis, and results of this exploratory temporal analysis using Landsat 8 and Sentinel-2 Remote sensing imagery, to classify land use, and explore land-use change in an area of Northern Virginia with a high saturation of data centers.

Map of the Remote Sensing Temporal Analysis Study Area (Northern Virginia)



Map 14

Study area covers roughly half of Loudoun and Fairfax Counties in Northern Virginia. The known data centers within the study area are mapped as red points. They represent 119 different data centers, operated by a variety of different companies. Most are located along VA State Route 267 in the towns of Sterling, Ashburn and Reston.

5.2.1 Data & Study Area

The study area (*Map 14*) for this analysis is an area in Northern Virginia containing parts of Loudoun and Fairfax counties, colloquially referred to as “Data Center Alley". Northern Virginia is the largest data center market in the US (St.Germain, 2019), and Loudoun and Fairfax counties have a particularly high density of data centers due to local tax incentives, and zoning policies (Dawn-Hiscox, 2020).[11] Due to the high number of data centers, this area is an ideal location to develop and test land-use classification methods, and conduct a preliminary land change analysis targeting data centers.

Several different datasets were used to perform this analysis and they include:

A Feature Collection of the 119 data centers that are located within the Study Area. I created this dataset using information contained in an online directory of data center locations (datacenters.com, n.d.). These data center location points are a subset of the Northern Virginia data center points used in the previous analysis. The data acquisition and cleaning process was the same for both datasets. Data was collected via the Python web crawler I created (*section 5.1.2*), geocoded, and QC'd to ensure that the data points were accurately geolocated. The data center points were then loaded into QGIS, transformed into a shapefile, and exported to be uploaded to the Google Earth Engine environment.

Remote Sensing data included the USGS Landsat 8 Surface Reflectance Tier 1 Image collection. Once loaded into Google Earth Engine, this image collection was filtered to the bounds of the study area, and delimited to the date range: (7/1/2015-9/30/2019) which resulted in a collection of 158 images which was utilized for the project analysis. The acquisition dates selected are targeting the same summer months to ensure that the vegetation is consistent across the composited images, and that no snow or other weather related interference is present in the images. The study area falls within two different Landsat scenes; the images are from the WRS Row/Path **015/033** & **016/033**.

---

[11] This study area represents a large portion of the Northern Virginia Study area from the analysis presented in the previous sections of this Chapter (1.3 & 1.4), refer back to those sections for more details about the spatial distribution of data centers within this area.

The Sentinel-2 MSI: MultiSpectral Instrument, Level-1C image collection was also used. This image collection was also filtered to the bounds of the Study Area and the same date range as above in Google Earth Engine, the resultant collection of 2,564 images was the dataset used for the project. These two remote sensing products were selected for this analysis for several reasons, first to test how the difference in resolution (30 meters for Landsat8 vs. 10 meters for Sentinel-2) between these products impacts the ability to accurately detect data center locations. These products were also chosen as they have a higher rate of return, and offer more abundant imagery to pull from than more selective remote sensing products such as NAIP imagery, which is collected via airplane at less frequent intervals. (NASA, n.d.)

5.2.2 Methods

Yearly Image Composites

The first step taken to prepare the image collections for further analysis was to mask the clouds. For the Landsat 8 collection, this was accomplished by creating a mask that used the Pixel QA band to select the pixels matching a particular QA band value. The cloud mask was then mapped over the Landsat 8 image collection using the *.map* GEE function. Similar steps were taken to prepare the Sentinel-2 collection as well. A cloud mask was created, targeting the QA60

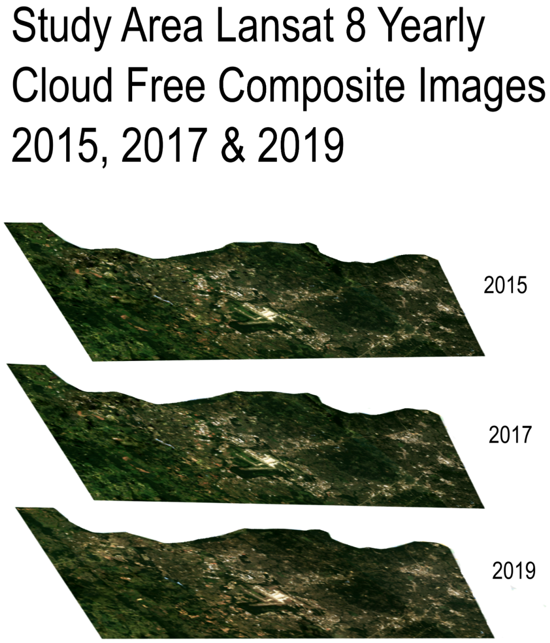


Figure 13: Landsat 8 cloud free yearly composite images of study area. Composites are for 2015, 2017, and 2019, and were made from imagery collected between (7/1 - 9/30) for each year. Compositing was done using the GEE mosaic function, and clouds were masked using the Landsat Pixel QA band.

band, using the *.bitwiseAnd* to select pixels with clear conditions.

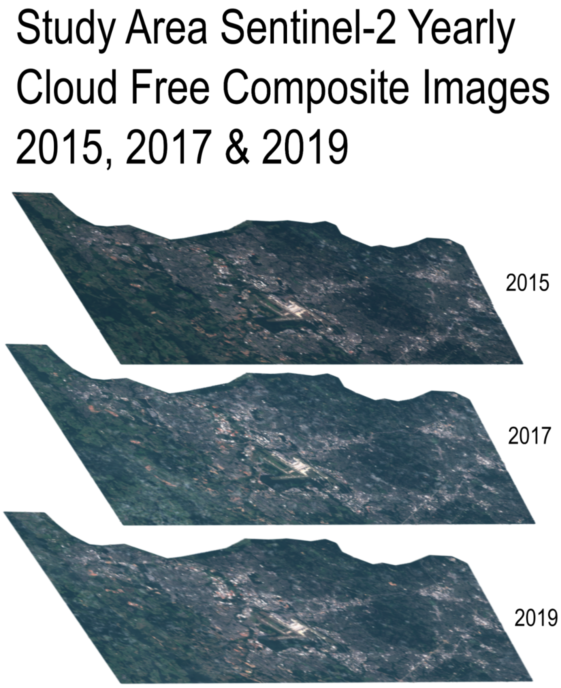


Figure 14: Sentinel-2 cloud free yearly composite images of study area. Composites are for 2015, 2017, and 2019, and were made from imagery collected between (7/1 - 9/30) for each year. Clouds were masked using the Sentinel-2 QA60 band and the GEE bitmask function.

This analysis will be looking in particular at the years: 2015, 2017, and 2019. After masking the clouds, composite images for each study year were created. Both the Landsat 8 and Sentinel-2 cloud masked image collections were filtered to the summer months (*7/01-9/30*) for each of the study years, and then these reduced image collections were mosaicked to create yearly cloud-free composite images (*Figure 13 & Figure 14*). As discussed, the composite images are all created from imagery collected during the same time of year and thus have a highly similar spectral signature which should increase the accuracy of the classification.

Supervised Classification

In order to detect changes in land use over the study area, the composite images needed to be classified. To accomplish this, a supervised classification of the composite images was done in Google Earth Engine, using the Random Forest Classifier (Breiman, 2001). The Random Forest classifier was trained using 245 different training points (*Figure 15*) representing six different land use classes: Water, Wooded Area, Field, Residential, Office/Retail & Warehouse/Data Center; these classes appear to be the dominant land cover for the study area. The classes were divided in this way for several reasons, namely, they were the most prominent

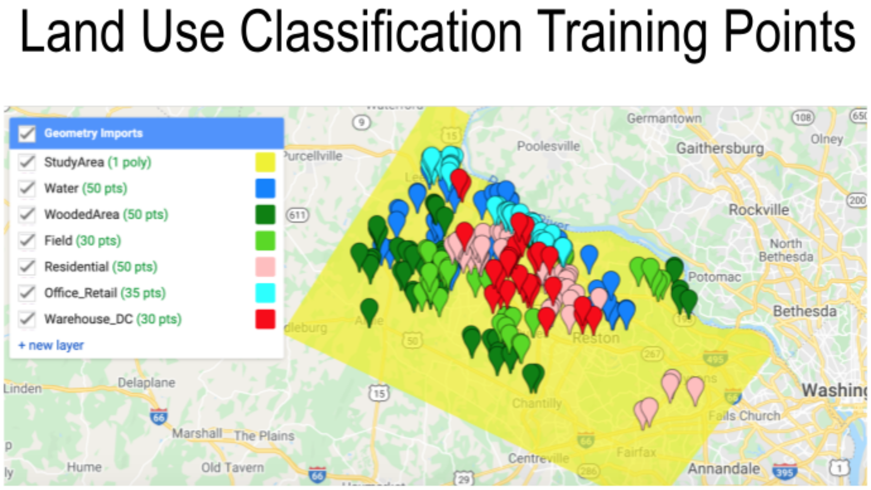


Figure 15: Map of the 245 training points representing 6 different classes used to train the classifier that was used to classify the Landsat 8 and Sentinel-2 Composite Images. The data center layer was not used in the creation of these points.

land uses from what I could discern from local zoning maps (loudoun.gov, n.d.), it seemed prudent to use a smaller number of classes as this was an exploratory approach without previous land classifications to rely on, additionally, as was borne out in the analysis from the previous section, data centers and warehouses/logistics buildings tend to be located near one another and have similar morphologies and visual signatures so grouping data centers and warehouses together was a decision to limit the potential for missed detection. Using the training points, and the composite images, training features were created, fed to the classifier, and then each composite image was classified into the six different land-use classes (*Figure 16 & Figure 17*).

A resubstitution error matrix was created for each of the classified images (*Figure 18 - Appendix B)* these were generated using Google Earth Engine's *errorMatrix()* function, which calculates the expected accuracy of the validation data that is created along with the training data when using the Random Forest Classifier (developers.google.com/earth-engine/classification). The error matrices show a high degree of training accuracy of around .99.

Study Area Lansat 8 Classified Images 2015, 2017 & 2019

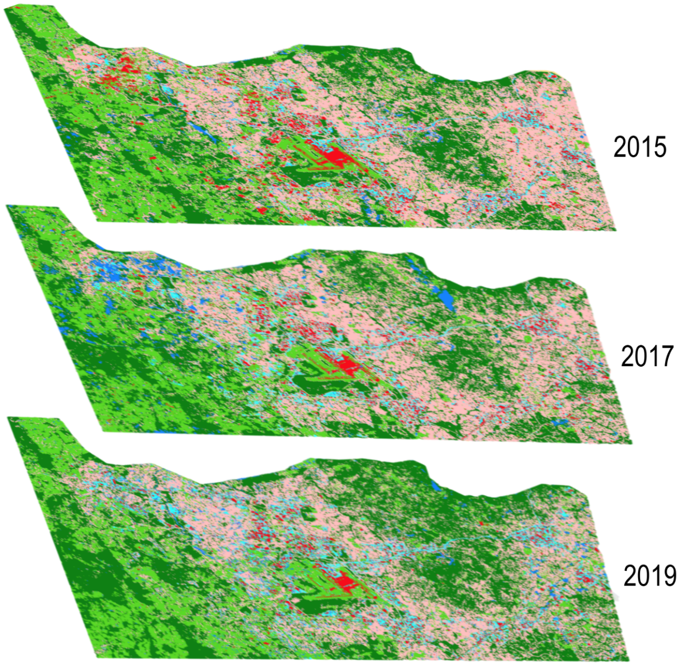


Figure 16: Classified Landsat 8 yearly composite images. Supervised classification using 245 training points representing 6 different Land Use / Land Cover classes. The 'Random Forest' Classifier, with 100 decision trees was used to create these images.

Group	*Name*	*Color*
0	Water	🟦
1	Wooded Area	🟩
2	Field	🟩
3	Residential	🟪
4	Office_Retail	🟦
5	Warehouse_DC	🟥

Study Area Sentinel-2 Classified Images 2015, 2017 & 2019

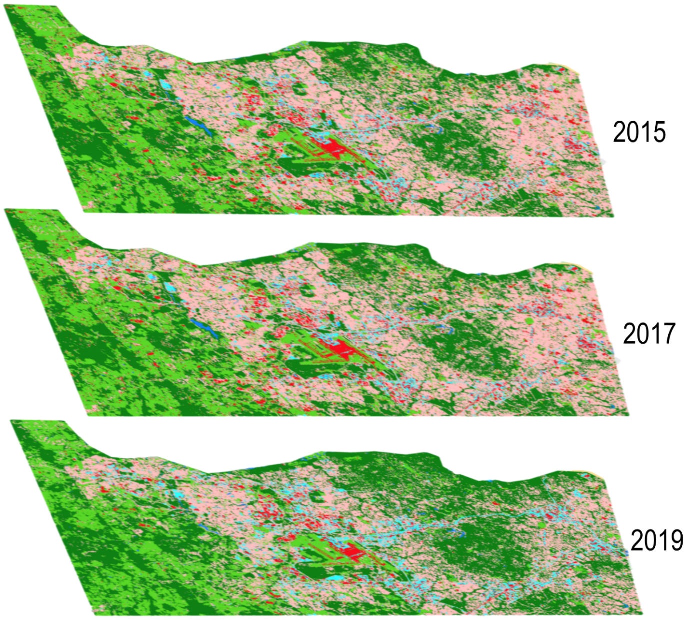


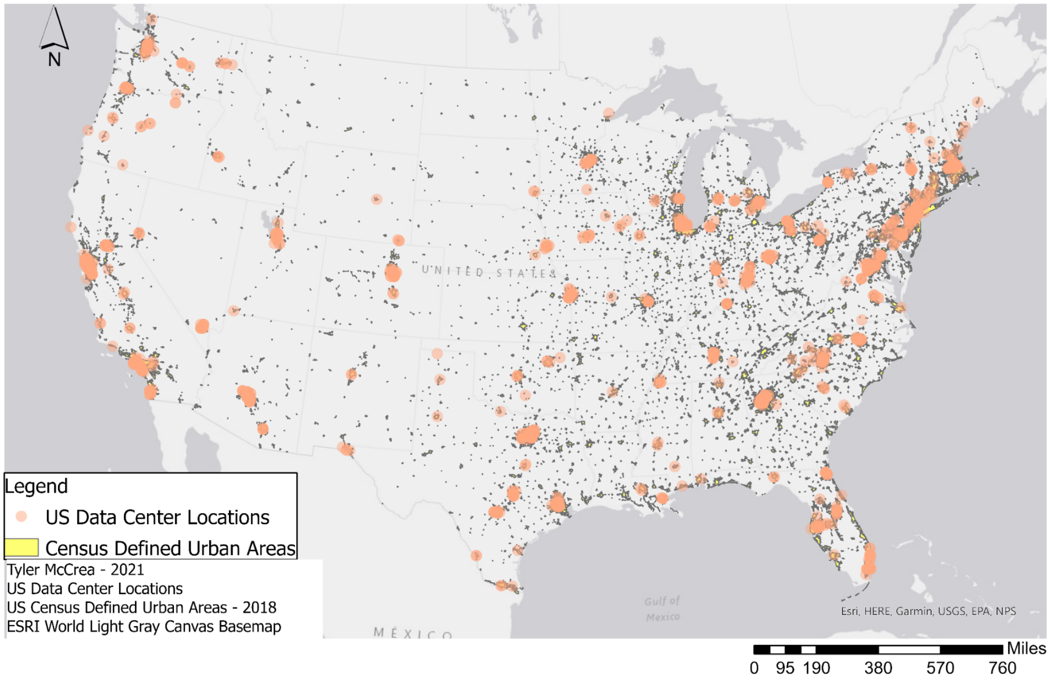
Figure 17: Classified Sentinel-2 yearly composite images. Supervised classification using 245 training points representing 6 different Land Use / Land Cover classes. The 'Random Forest' Classifier, with 100 decision trees was used to create these images.

Group	*Name*	*Color*
0	Water	🟦
1	Wooded Area	🟩
2	Field	🟩
3	Residential	🟪
4	Office_Retail	🟦
5	Warehouse_DC	🟥

6. Results

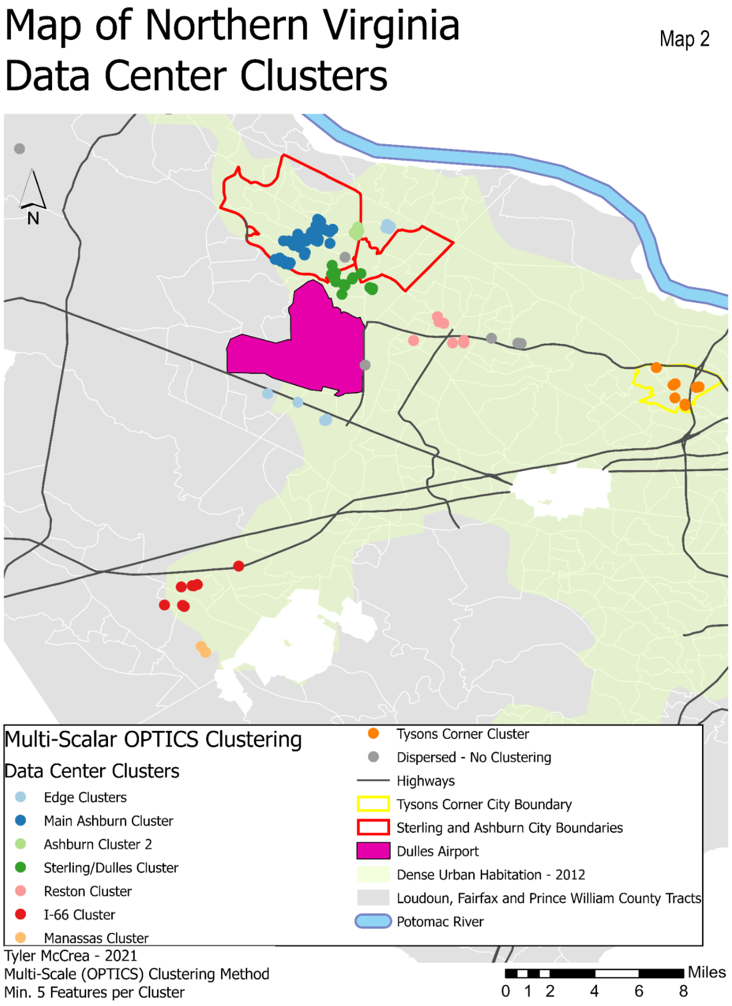
Map of US Data Centers and Census Defined Urban Areas

Map 1



Before diving into the distribution and clustering analysis of data centers in the study area, it seems pertinent to begin with the map above (Map 1), which shows the location of approximately 1,300 data centers across the US. From the map, we can see that the distribution of data centers appears to be highly clustered in and around large and mid-size cities. More than 95% of these data centers are located within a census defined urban area and bears out the research that data centers tend to be centralized around existing infrastructure, but contradicts some recent scholarship which frames data centers as being rooted in rural areas for ‘urban’ data extraction and processing (Levenda & Mahmoudi, 2019). There are certainly examples of hyperscale data centers in largely rural areas, but the vast majority of data centers are located in and around cities, and in general target areas with significant tax and financial incentives (NVTC, 2020).

6.1 *Spatial Patterns*



Tyler McCrea - 2021
Multi-Scale (OPTICS) Clustering Method
Min. 5 Features per Cluster

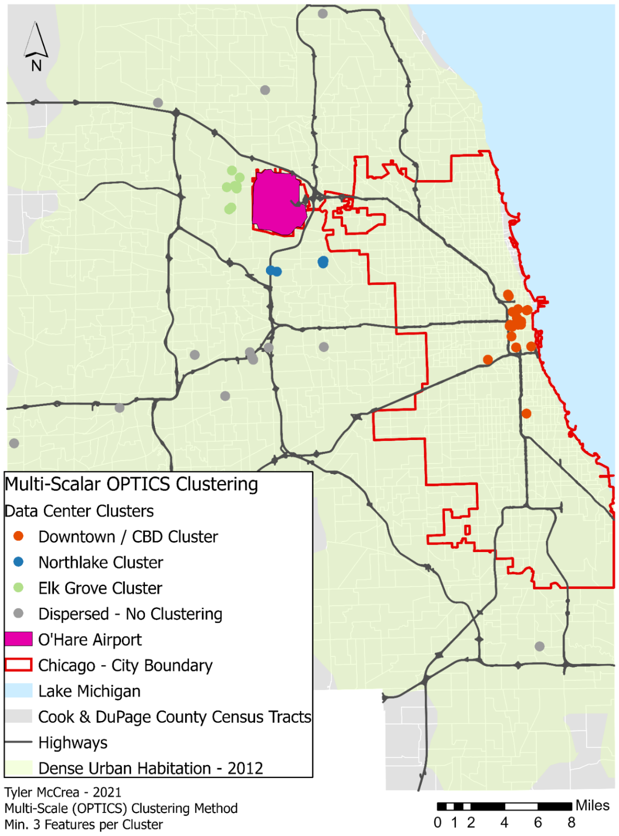
The first target of this spatial analysis is the spatial distribution of data centers across the study areas, as discussed in the methods section, this was accomplished by running a Multi-scale OPTICS clustering algorithm over the data center point data. The output of the OPTICS cluster analysis can be seen in the reachability charts (Figure 4, Figure 5 & Figure 6), there is a clear distinction between the various study areas, foremost in the number of points in each plot. The Northern Virginia study area (Figure 5) has twice as many data center locations as the Chicago or Phoenix area plots (Figures 5 & 6). Visually interpreting the data on the plots there is also a fairly clear distinction between the density of the clusters present in the study areas. The Northern Virginia, and Chicago Study areas have denser clusters, characterized by steep valleys between the first and last data points of a cluster, whereas Phoenix only shows one highly dense cluster, exhibiting more of a dispersed characteristic than the other study areas. Also interesting to note is the number of clusters for each study area, Northern Virginia, in addition to having dense clusters, also has the greatest number of clusters, whereas the Chicago area has the fewest of all the study areas with only three sets of clusters.

In addition to the reachability plots, the OPTICS clustering tool in ArcGIS Pro also outputs a Shapefile layer with the points classified according to their cluster, as determined by the reachability plots. I used these cluster layers to produce maps of each study area, showing the clusters along with regional highways/interstates and local airports, to provide context, and to depict data centers in relation to other pieces of infrastructure. Looking first at the Northern VA map (*Map 2*) there is a clear association between the location of data centers and highways, as well as data centers and the airport. The largest cluster in the Northern VA study area with 66 data center locations (*Figure 1 - Appendix B*) is located within the town of Reston, VA, an area of data center saturation known for its adjacency to state intelligence and military installations, as well as an area with targeted economic incentives for data centers in the form of tax abatements.

The trend of data centers clustering based on (1) tax incentives, (2) near highways, and (3) around airports holds true and is evidenced across all of the study areas. The highway and airport adjacency trend is particularly evident in the Chicago (*Map 3*) and Phoenix (*Map 4*) metropolitan study areas, nearly every data center in both of those study areas is within .75 miles of a highway, only a few outlier points are further afield. Additionally, in both Chicago and Phoenix study areas two or more prominent clusters exist in the area around the major airports. The trend of data center locations falling along highways, and near airports is a curious one, and follows a similar pattern to that of logistics infrastructure, specifically fulfillment and distribution centers (Sheffi et al., 2019; Good Jobs First, 2021).

Map of Chicago Area Data Center Clusters

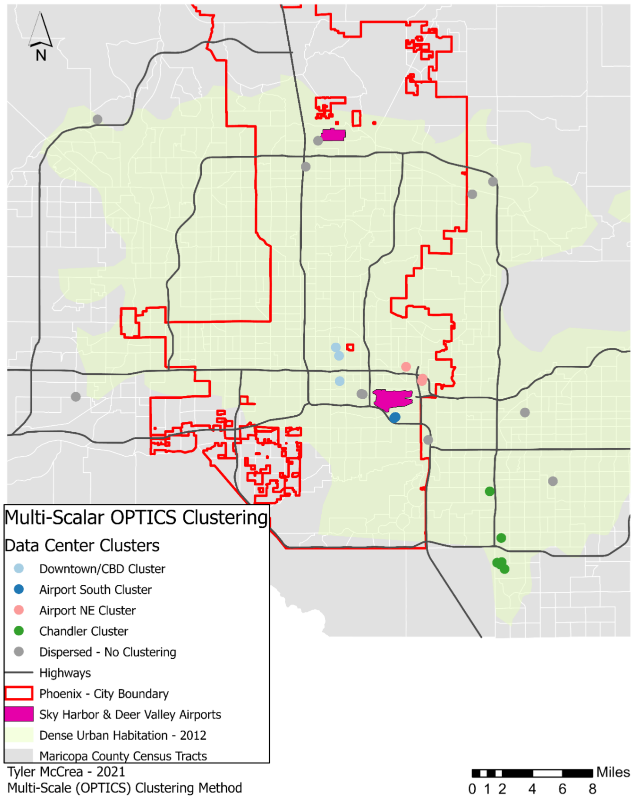
Map 3



Tyler McCrea - 2021
Multi-Scale (OPTICS) Clustering Method
Min. 3 Features per Cluster

Map of Phoenix Area Data Center Clusters

Map 4



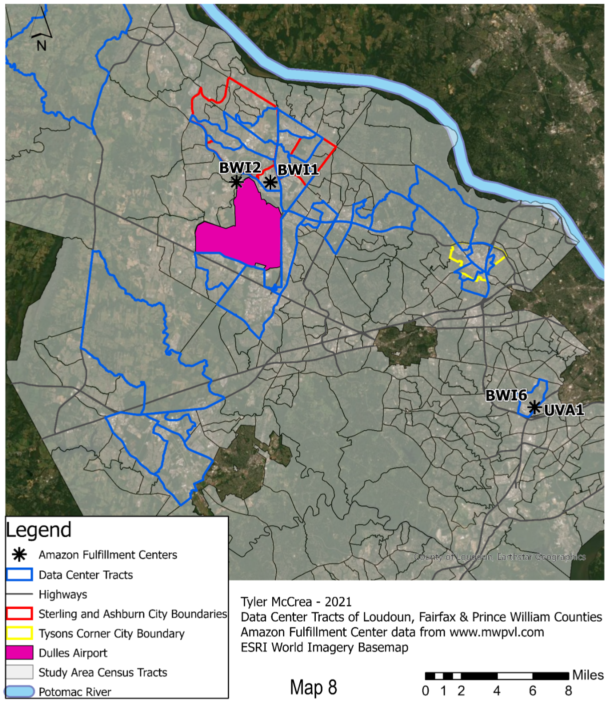
Tyler McCrea - 2021
Multi-Scale (OPTICS) Clustering Method
Min. 3 Features per Cluster

In a recent report from the non-profit Good Jobs First, they conducted a spatial analysis of the locations of amazon fulfillment centers, mapping their construction patterns over time, and drawing out high-level trends about the patterns they saw. The main determinants for the siting of fulfillment centers were: proximity to highways, airports, and adjacency to urban areas (Good Jobs First, 2021). Access to air and ground transportation makes sense as a priority for this infrastructure as physical goods need to travel by air or road. What's curious is the overlap of these characteristics with data centers, which ostensibly do not require trucks or planes to go about their work of collecting, processing, and storing data. On the other hand, the collection of these infrastructures near one another makes sense as they're mutually constitutive, the modern fulfillment center would not function without the software integration provided by the data center, and data centers have grown precisely because of data-intensive changes to industries. In this way, the data center/fulfillment center exemplifies the idea of 'code/space', spaces which are produced or (re)produced by the dyadic relationship between software and physical space (Kitchin & Dodge, 2011).

In order to better understand the association between the clusters of data centers and the locations of Amazon Fulfillment centers, I compared the distances between these two infrastructure objects across the study areas. First, I aggregated the data center locations to their underlying census tracts (*Map 5, Map 6, Map 7 - Appendix A*), and from this, I was able to identify all of the 'data center tracts', any census tract with one or more data center locations. Having identified which tracts containing data centers, I mapped these tracts along with Amazon Fulfillment Center locations (*Map 8, Map 9, Map 10*) for each study area, and measured the distance between each fulfillment center, and the nearest data center tract (*Table 1*).

In all study areas, we can see a high agreement between the locations of the data center census tracts and the location of fulfillment centers, which are on average less than 1.5 miles from each other. Amazon Fulfillment centers in the Chicago and Northern Virginia study areas were particularly associated with data center census tracts, barring one outlier, every fulfillment center was located within a data center tract.

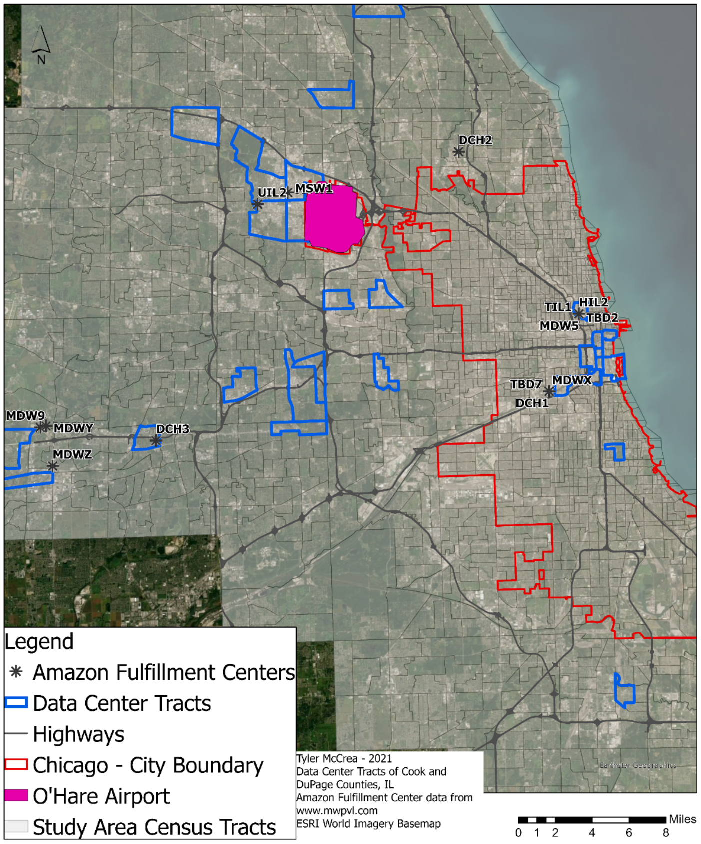
Map of Northern Virginia Data Center Tracts & Amazon Fulfillment Centers



Tyler McCrea - 2021
Data Center Tracts of Loudoun, Fairfax & Prince William Counties
Amazon Fulfillment Center data from www.mwpvl.com
ESRI World Imagery Basemap

Map 8

Map of Chicago Area Data Center Tracts & Amazon Fulfillment Centers



Tyler McCrea - 2021
Data Center Tracts of Cook and DuPage Counties, IL
Amazon Fulfillment Center data from www.mwpvl.com
ESRI World Imagery Basemap

Map 9

Distance Table			
Distance from Fulfillment Center to Data Center Tract	*N. VA Study Area*	*Chicago Study Area*	*Phoenix Study Area*
Mean Distance (mi.)	0.2	1.1	1.6
Median Distance (mi.)	0	0.2	1.2
Max Distance (mi.)	0.8	11.1	3.1
Minimum Distance (mi.)	0	0	0

Table 1 - This table depicts the summary statistics for the simple distance based analysis, measuring the distance between each Amazon fulfillment center and the nearest data center census tract.

The association between highways, airports is the most apparent across all study areas when viewing the cluster maps, but also evidenced in the presence of data centers in the edge or industrial areas, as well as clustering in and around central business districts. The clearest example of this is in the Chicago study area (*Map 3*), where the largest cluster of data centers is in or near the ‘Loop’ of the central commercial and financial district, there are small CBD clusters in Phoenix, as well as in Northern VA near the small exurban commercial hub of Tysons Corner. The clustering near logistical and financial infrastructure and customers provides evidence to support the assertions made at the beginning of this research paper that at their core, data centers are sites of value production and profit-seeking, and their locations bear that out

6.2 *Place-based Analysis*

Image 1: Data centers which are examples of the faceless building typology (type 1)

After researching cloud computing and data centers over the course of this project, I started seeing many targeted ads trying to sell me cloud computing resources, or colocation services. Many of these advertisements depict data centers as crisp blue-hued rooms of servers or pure white rooms akin to the clean rooms used in microchip productions (Taylor, 2017). Several geographers and other scholars have also written about the presentation of the cloud and data centers as a ‘people-less techno-wilderness’ (Taylor, 2019), or considered the ways that certain data center designs embody a sort of power-laden logic in their aesthetics (Laparelli, 2020). The connection between aesthetics and logics of cloud computing is an interesting aspect to explore, this brief section considers the placial characteristics of the cloud.

Configuring what the average data center in each cluster actually looks like, not via coordinated marketing materials but in the 'real world'[12].

Image 2: Data centers which are examples of the office park building typology (type 2)

Using the data center clusters that were identified in the previous section, I randomly selected one data center from each cluster, and visited the location using Google Street View. I took a screenshot of the data center building and then collected all of the photos to compare the different morphological characteristics for the buildings. Focusing particularly on the building layout, aesthetics, and orientation in relation to the built environment. Based on the characteristics I found, I identified three different classifications for building types: Type 1 - Faceless; Type 2 - Faux Office Park; Type 3 - Reuse. These types were evidenced across clusters and study areas.

The most prominent of these types, in appearance and number, the *Faceless* building type identified above (*Image 1*), data centers of this type are generally very large structures, with few to no windows. These structures are imposing, with vast blank walls with few distinguishing details, and are usually surrounded by gates or fencing. This type of data center is usually found in edge zones or within industrial parks, particularly noticeable in the clusters around airports.

---

[12] Using quotes here since technically these sites were 'visited' via google streetview due to the varied locations, and travel restrictions due to the global pandemic.

The next most prominent building type was what I classified as the ‘Faux office park’ building type, which shares many characteristics with ‘faceless’ data centers, but instead of large empty walls, these centers are built with ornamentation and a large amount of mirrored glass. In function these data centers operate exactly as those of the previous type, the addition of the glass is purely an aesthetic decision, which gets at the idea of shaping public perception around this infrastructure, moving from hidden to ‘open’.


Image 3: Data centers which are examples of the adaptive reuse building typology (type 3)

The final building type, which was the least prevalent, was the reuse building type, where a data center is constructed in a building that was previously used for a different purpose. Although this style was most common in the Chicago study area, examples could be found from each study area, and this type is usually located in or near the central business district, and separate from the clusters near airports.

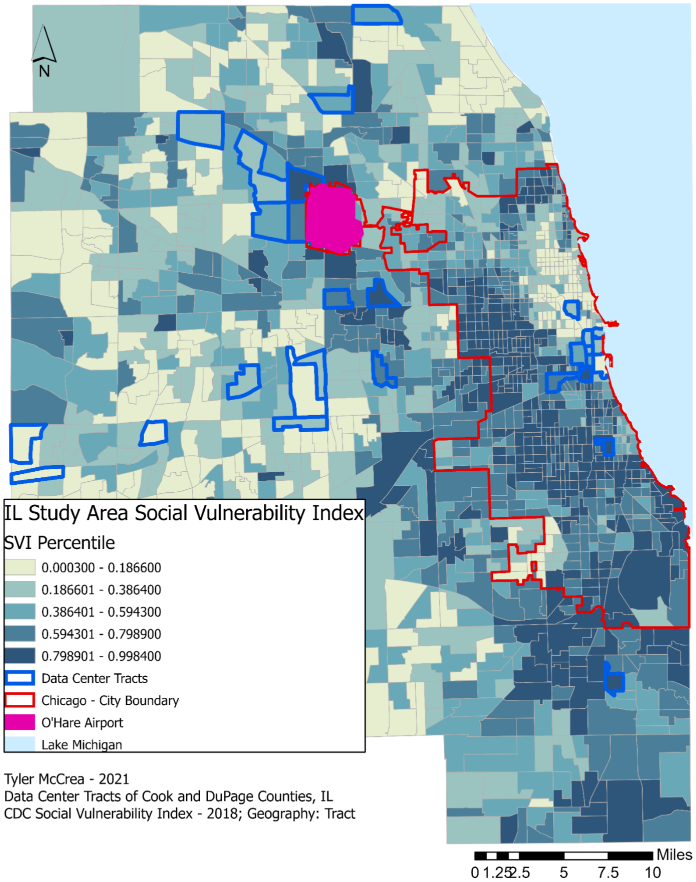
The general appearance and characteristics of the data center building is of interest to geographers, and in particular human geographers due to the relationship between the built environment and the ‘sense of place’ (Nelson et al., 2020) . Yi-Fu Tuan (1977) wrote that place is space made meaningful. When considering the building types described above, the placed-ness of this infrastructure is driven by multiple logics of concealment in the faceless building type, and in some instances presenting an outward appearance that belies the true nature of the building, and its functions.

6.3 *Community Characteristics*

In addition to investigating where and in what forms data centers can be found and mapping their distribution and position, I am also curious about the general composition of the communities of people who live around data centers. I am not interested in drawing any conclusions, based solely on the demographics, or community characteristics of communities around data centers, but rather I am seeking to continue the work of previous scholars in this field by exploring whether or not specific the dynamics between data centers construction and community health, demographics, etc. can be elucidated through spatial analysis. In order to explore this dynamic, I relied on a ready-made dataset by the US Centers for Disease Control which assesses community vulnerability at the census tract level using four main variables to assess vulnerability: socio-economic status, household composition, minority status,

Map of Chicago Area Data Center Tracts & Social Vulnerability

Map 12



Tyler McCrea - 2021
Data Center Tracts of Cook and DuPage Counties, IL
CDC Social Vulnerability Index - 2018; Geography: Tract

and housing type (CDC, 2021). Communities with a high prevalence of households without adequate transportation or with high rates of unemployment would have higher SVI values than communities with higher incomes.

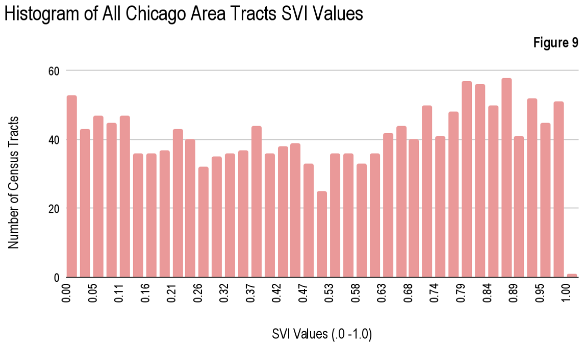


Figure 9

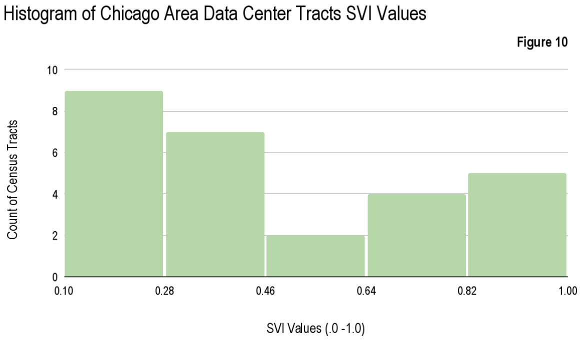
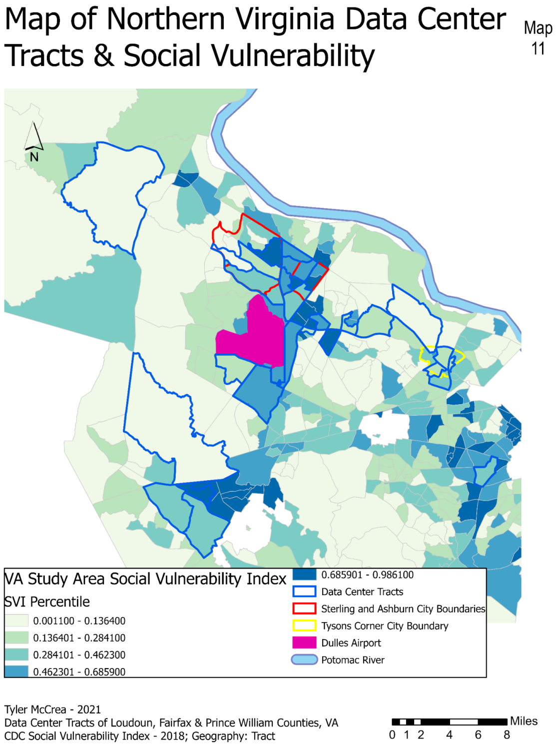


Figure 10

At the start of this analysis I had assumed that there would be a high correlation between data center tracts (*see section 5.1.2*) and census tracts with high social vulnerability scores (in relation to the study area as a whole). However, the analysis has borne out that on average the data center census tracts in each study area actually have lower SVI values than the study areas as a whole. This is particularly evident in the Chicago study area SVI map (*Map 12*) and the SVI histograms (*Figure 9 & Figure 10*) despite the overall study area having a median SVI score of '.5355' the median SVI across only the data center tracts was



Map 11

Tyler McCrea - 2021
Data Center Tracts of Loudoun, Fairfax & Prince William Counties, VA
CDC Social Vulnerability Index - 2018; Geography: Tract

‘.3813’. Similar trends were observed across all of the study areas, albeit less pronounced than the stark difference evident in the Chicago study area. For example in the more prosperous on less vulnerable Northern Virginia Study area (*Map 11*), the data center tracts had a slightly lower SVI value than the average for all tracts, but interestingly the data center tracts tended to neighbor areas of relative vulnerability.

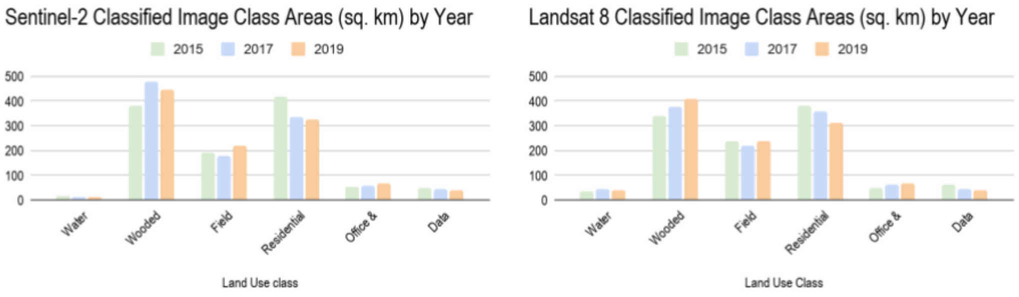
While this analysis is not meant to draw any definitive conclusions from these associations, it is interesting to see how the data center tracts tend to have lower social vulnerability than the study areas as a whole, and also correlates with one of the findings from the ‘Mapping Amazon Project’ discussed previously, which is that fulfillment centers tend to be located near areas of relative prosperity, which seems to hold true for data centers as well.

6.4 ***Temporal land-use change***

The next section considers the results from the remote sensing-based land-use change temporal analysis.

Change in Class Area

Along with visually comparing the yearly classified images derived from Landsat 8 and Sentinel-2. I also want to look at specific changes in land use over time, in order to understand how data centers fit into larger land use trends, and


Sentinel-2 Classified Image Class Areas (sq. km) by Year

Landsat 8 Classified Image Class Areas (sq. km) by Year

Figure 21 - Land use class areas for both Sentinel-2 and Landsat 8 classified composite images. Charts show the class area in square kilometers for each year of the analysis. Class area statistics were exported from Google Earth Engine.

understand how the data centers have trended over time. A fairly simple way to visualize these trends, and also to examine the differences between the Landsat 8 and Sentinel-2 classified images, is to depict the changes in area by class for each of the yearly classified images. Using the dictionary function to calculate the area of each class, two charts were generated using data from the classified images (*Figure 21*). The charts show the area of each of the 6 classes in square kilometers, by year for the Landsat 8 and Sentinel-2 classified composite images.

The class area charts show a fairly symmetrical trend in the Landsat 8 yearly class areas, with more variability in the Sentinel-2 classified images from year to year. The most significant fluctuation for the study dates was in the wooded and residential classes, which I would attribute to the low-density nature of many residential areas, an abundance of street tree canopy which makes the residential class hard to distinguish at times. When looking at the Office/Retail and Warehouse/Data Center classes in Figure 21, there are similar trends in the Landsat 8 and Sentinel-2 classified images, a slight decrease in Warehouse/Data Center and a slight increase in Office/Retail. This would seem counter to the prevailing development trends, but upon further inspection of the classified images, the Office/Retail class was classifying mostly parking lots and paved surfaces, as well as dark-colored flat-roofed structures, and classified more parking lots in 2019 vs. 2015.

Dulles International Airport is located within the study area, had a large

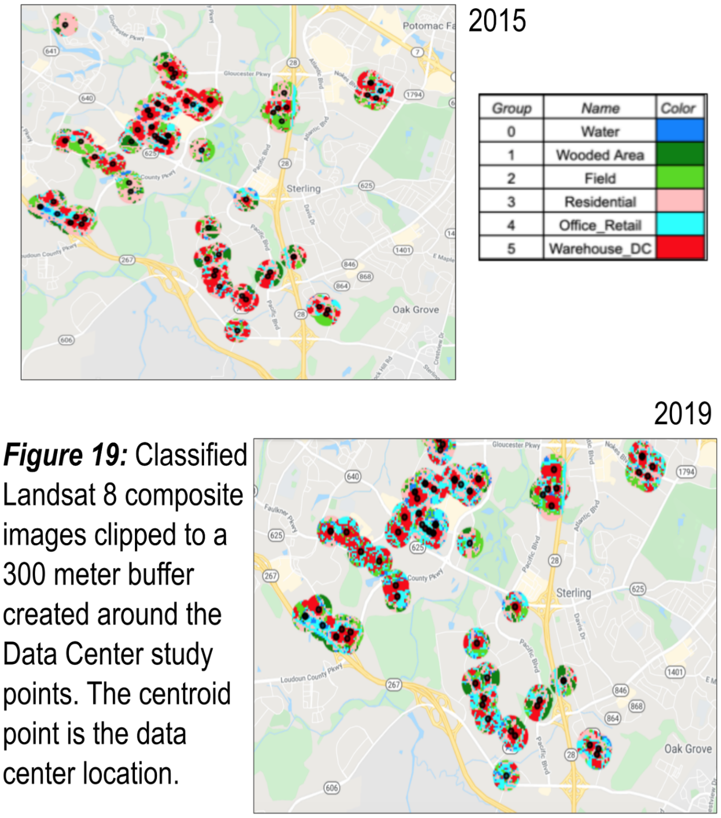
Figure 22 - Land use class areas for both Sentinel-2 and Landsat 8 classified composite images with the area of Dulles Airport masked to remove distortion. Charts show the class area in square kilometers for each year of the analysis for the Office_Retail and Warehouse_DC land use classes.

portion of its land area classified as a Warehouse/Data Center, which significantly added to the area for this class, but also distorted the class as the airport clearly wouldn't fall under either category. In order to mitigate some of this distortion the land area was recalculated, this time with the area covered by Dulles Airport masked (i.e. removed). In the charts below (*Figure 22)* with the distortion removed from the airport class we can see the data center development trends more clearly in the data, with the exception of the Landsat 2015 classification which seems to be an outlier.

Data Center Accuracy

One aspect of the temporal analysis was also to determine how effective the land use classifier is at detecting data centers. As such, one additional method of analysis will be to look at the classified image outputs in relation to known data centers. To do this I first needed to transform the data center feature collection into a single feature. In Google Earth Engine you cannot buffer a feature collection, so the points needed to be reduced to a single feature. Using this new

Classified Landsat 8 Images Clipped to Data Center Buffers

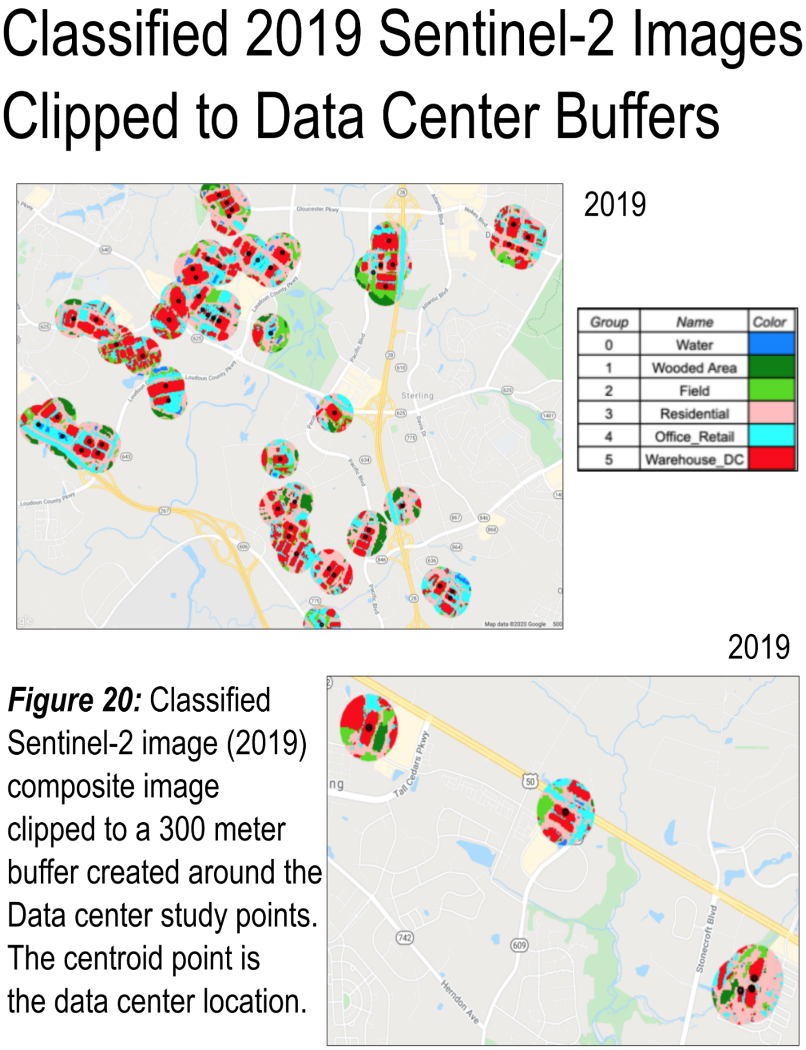


Group	Name	Color
0	Water	🟦
1	Wooded Area	🟩
2	Field	🟩
3	Residential	▨
4	Office_Retail	🟦
5	Warehouse_DC	🟥

Figure 19: Classified Landsat 8 composite images clipped to a 300 meter buffer created around the Data Center study points. The centroid point is the data center location.

feature, the *.buffer* function was run to create a buffer of 300 meters around each data center location. The buffer radius was chosen based on the average size for a data center, and the desire to target and see the accuracy of the classifier at a fine scale. This feature was used to create maps showing just the area of the classified images within the buffer zones *(Figure 19* & *Figure 20*) discussed in more detail below.

To quantitatively test the accuracy of the classifier in detecting data centers further analysis was performed using the Data Center study buffer feature discussed in the Methods section. Each of the yearly classified images was clipped to the Data Center buffers and each of the data center points was added to determine if the classified images were correctly classifying data center pixels in the Data Center class. The Landsat 8 classified image clipped to the study points (*Figure 19*) shows a fair degree of accuracy for 2015 and 2019. Some data centers that were classified correctly in 2019 are classified as a different land class in 2015,



Group	Name	Color
0	Water	🟦
1	Wooded Area	🟩
2	Field	🟩
3	Residential	🟪
4	Office_Retail	🟦
5	Warehouse_DC	🟥

Figure 20: Classified Sentinel-2 image (2019) composite image clipped to a 300 meter buffer created around the Data center study points. The centroid point is the data center location.
.

providing a look at land tenure change specifically related to data center construction. Observing this land-use change is expressly the sort of material transition I was hoping to discern and adds credence to this being an effective method for getting at the mixed and messy nature of data center materiality.

Sentinel-2 classified images were also clipped to the study points (*Figure 20*), and appear to be highly accurate. Focused on a large cluster of data centers as well as an outlying smaller group, the images show a high agreement between the classified image and the actual data center location, even providing enough detail to see the building outlines. Observing the Sentinel-2 data center buffers between years, we can also see evidence of data center locations transitioning from one land-use class to another over time, and using Sentinel-2 data produced more accurate representations of that change. The difference here is largely due to the improved resolution of Sentinel-2 data which is 2x as refined as the Landsat8 Imagery.

Using this clipped feature, we can assess the percentage of data centers that were correctly identified for each classified image product, and data center land-use trends between 2015-2019. To assess the accuracy further I created a table (*Table 3 - Appendix C*), which depicts for each of the 119 data centers, their class for each year of analysis for the Landsat 8 and Sentinel-2 classified image composites. Comparing the 2019 classifier outputs for Landsat 8 & Sentinel-2, the Landsat 8 classified image had an agreement between 83 of 119 data center locations, meaning 69.7% of the known data centers were correctly classified in the image. The Sentinel-2 classified image for the same year had a success rate of 84.8% or 101 of 119 data centers.

7. Discussion

7.1 *Spatial & Place-based Approaches*

The methods of spatial analysis described above are an attempt to work on the research gap identified by Pickren (2018), that locating, mapping, and spatializing the data center remains a key task in identifying and revealing the materiality of data centers. While others have examined specific hyperscale data centers in rural areas (Levenda & Mahmoudi, 2019), or the conversion of military infrastructure to data centers (Fish & Garrett, 2019), this work considers the typical data centers as they exist in various metropolitan areas through their spatial distribution, form, and community characteristics.

The first aspect of this analysis, finding and processing the location data for data centers was far and away the most difficult, and time-consuming portion of this work. Peeling back the layered (in)visibilities (Furlong, 2020) of location, size, power consumption, and every other scattered data point proved impossible. As I discussed in the data collection section above, the fixation on constructing the ‘perfect’ dataset became reductive and ended up prioritizing finite locations over material understandings. Despite the issues encountered during data collection, or perhaps because of them, it’s clear to see that despite a relative “opening up” of data centers in many ways these infrastructures are still hidden and obscured. While the internet may be considered “open” or “democratic”, much of what comprises these open networks is the highly privatized space of the data center, and its related infrastructure encompassed in the private telecommunications networks.

This analysis has borne out that data centers within the US are clustered primarily around existing financial and commercial hubs and oriented around existing “non-digital” infrastructure. Additionally, at the local level of the three study areas, the analysis has shown that data centers tend to cluster in several parts of a given metropolitan area, usually in or around the central business district, near airports, and within industrial and office parks, usually along major highways. This is to say, that they are positioned and deployed with the primary function of maximizing profitability for the data center operator, dependent upon existing infrastructure, and clustered in areas with tax and financial incentives. They exist not to act as a bastion of the free and open internet of the imaginary, but rather to serve the purposes of global capital. To maximize labor time and minimize production time, to process and valorize data through commodification (Sadowski, 2020).

Again, it needs to be stressed that these methods were deployed primarily for an investigatory analysis of the nature of data centers and to explore ways of attending to the need to spatialize the data center as well as uncover the hidden materialities of data centers. These methods succeeded in that I was able to construct a comprehensive dataset of data center locations across several metropolitan areas, compare the presence and form of data centers in those areas, and present a well-rounded picture of the nature of data centers in those areas. Where these methods fall short is that I was not able to address the total impact of the operation of these data centers without accurate information regarding the size and energy consumption patterns of the data centers. Additionally, without additional datasets, and more stringent statistical methodologies, a more definitive assessment of the material and social impacts of data centers is impossible. In the following sections, I attempt to circumvent the issue of lack of data, and explore the use of remote sensing to ‘sense’ the cloud, and examine its material conditions.

7.2 *Remote Sensing Approach*

The results of the temporal land-use land-change analysis, as preliminary as they are, point to Sentinel-2 imagery as proving more accurate for classified image analysis, this is largely due to the higher spatial resolution which allows for the classifier to more accurately classify pixels, particularly on the edges between classes. Where Landsat 8 class boundaries tend to be chunky in appearance, the Sentinel-2 classified image boundaries were smooth and often matched existing building outlines. From the data center accuracy test we can also definitively say that the Sentinel-2 classified images had a greater degree of accuracy at known data center points vs. Landsat 8. We can also see some preliminary results in looking at which land classes are the most common for data centers to be built upon. In Table 1 we can see data centers that are correctly identified in the data center class in 2019 but are in a different class in 2015 or 2017. The most common previous class for a data center is the "Residential" class, which seems somewhat unlikely, but could be due to the classifier confusion between residential and wooded areas, and the fact that residential areas and data center zones are often adjacent to each other within the study area.

The intention of this analysis was to determine if remote sensing data was a useful tool for conducting land use and land change analysis related to data centers in a mixed-Exurban environment. I believe it was a success in that I was able to experiment with two different products, compare their outputs, and conduct a basic analysis, and accuracy test. I believe with more classes, and training points, supervised classification could be an effective method to conduct temporal analysis, and to identify data centers among other building structures. Without more stringent verification of results, however, this method is susceptible to a high degree of error. It would be most effective when combined with additional data sources such as city and county tax parcel information. In the future, I would like to expand upon this research and use these methods to conduct an analysis in multiple regions to further explore regional trends, but also to assess the accuracy of these methods in built environments with varied morphologies.

8. Part B Conclusion & Next Steps

Looping back to the research questions which guide this chapter, the two spatial analysis methods detailed above were oriented to examine what could be gleaned from a spatial analysis of data center locations, as well as to test the feasibility and effectiveness of spatial approaches for a materialist analysis of data centers. To that end, the spatial analysis of data center distribution and building type across study areas, as well as temporal land-use change analysis which were undertaken, have elucidated where data centers tend to be clustered, the dominant building types, as well as revealed data centers, to be a distinct land cover type along with associated infrastructures such as warehouses. It must be noted that this analysis was primarily focused on testing these methods for further study and so specific attention was paid to the positives and negatives for the approaches, along with consideration for further research.

Beginning with the GIS-focused spatial analysis (Section 1.3), these methods allowed for an analysis that was able to provide a more complete picture of the ***where*** and ***what*** of data centers across different areas. Allowing for analysis related to distribution and the incorporation of additional social datasets. Unfortunately, this analysis relies heavily on data collected from private industry sources which are hard to both obtain and verify. Simply plotting the locations of data centers on a map doesn't inherently prove anything, but mapping is important to add a spatial aspect to the study of infrastructure and is crucial to a materialist analysis.

With regards to the remote sensing based methods (Section 1.4), this approach yielded positive results in terms of accuracy of detecting data centers and allowing for a robust analysis without specifically relying upon external data sources, negating the confounding layered (in)visibilities that were discussed in the data collection section. Remote sensing also incorporates inherently materialist framing as we are examining change to the natural and built environment. This approach is not without its flaws, first, a more all-encompassing study would still require additional data sources for quality control and validation, and second, this approach when applied only to the data center does not allow us to study the material impacts of data centers beyond their physical location.

The spatial and temporal analyses which were reviewed in this chapter were conducted not to draw out definitive conclusions, but to serve as the basis for further analysis, while building upon existing work, and attending to research gaps around data centers identified by geographers and others. While there are definitely positive outcomes and usefully exploratory results, these methods are quite preliminary, and undeveloped in their extent, but can be built upon in a way that will further elucidate the material and social dynamics of data centers. Primary to further study and application of these methods is the development of a theoretical model of critique which will frame the future work and determine what aspects of these methods to carry over. I attend to the work of synthesizing a theoretical understanding of data centers in Part C, formulating a conceptualization of data centers that will allow for more informed and robust studies in the future that integrate the successful aspects of the methods tested here, which lack a refined socio-political framework.

Part C - Space, Time, Extraction & Accumulation: Toward a Conceptualization of Data Centers as Multi-Scalar Nexus of Technocapitalist Relations

Part C Abstract

This part of the thesis focuses on developing a conceptual foundation for research around data centers, synthesizing and adapting existing theories to understand the sociotechnical, political, and material impacts of data centers. I locate this work within the broader Geography subfield of Digital Geographies and review the approaches and methods encompassed. I argue for a new conceptual approach to understanding data centers' role and their position within systemic flows of capital and resources using theories such as Technocapitalism and Granular Geographies. These theories ground the proliferation of data centers and cloud computing within the systems of capital production, and situate the data center as a spatial object with unique material and more than material impacts across multiple scales. I conclude by laying out future research directions and note the limitations and questions that remain.

9. What's missing *or* Why Theorize?

The previous section of this document (*Part B*) focused on an initial investigation into data centers using GIS and remote sensing methods. The spatial analysis in Part B was rooted in a materialist framework, and as a result of that analysis, I was able to draw some key conclusions about the logic and nature of data centers. Namely, that their distribution was highly clustered along highways, near airports, and contingent upon tax and economic incentives. Based on these results, and critical research regarding the nature of technology and society (*see Literature Review - Section 2*) I posited that the spatial logic of data centers at the multiple scales they were observed is driven by capital and financial interests, rather than a distributed "open" internet. One of the main conclusions drawn from the analysis in Part B is that, in order to more deeply attend to the material, socio-technical and political impacts, and dynamic of data centers, further theoretical frameworks were needed to understand how this infrastructure operates at multiple scales, and to decompose its many-layered materialities (Furlong, 2020). This section is oriented towards a conceptual and theoretical approach. I present three cross-disciplinary theories, synthesize them, and adapt them in tandem to both situate and conceptualize the data center as a site of multi-scalar social, economic, political, and material relations.

Following this introduction, I will begin by situating this work within the broader geographic subfield of Digital Geographies, scholarship that treats the 'digital' as both the subject and object of inquiry (Ash et al., 2018). Adding to the larger literature review at the beginning of this paper, which lays out the main research foundations from the broad umbrella of Science and Technology Studies to the narrow focus of Critical Data Center Research. This digital geographies overview is meant to explore some of the existing theoretical and methodological approaches geographers employ to understand “the digital”, and the myriad interrelated, political, economic, and social impacts. As well as to situate this research within this field, and emphasize the importance of further data center research, and broader critical-material inquiry around digital infrastructures of all kinds.

Following the Digital Geographies review, I lay out a theoretical framework for conceptualizing data centers as political objects with unique spatiality and power that are operationalized to extract and accumulate both data and capital, and also dependent upon, and enabling wide-scale resource and material extraction. To do this I will draw upon several different theories, namely: Technocapitalism (Suarez-Villa, 2009; 2016), the Planetary Mine (Arboleda, 2019; 2020), and Granular Geographies (Jamieson, 2020) and discuss how these ideas, when applied to data centers, can begin to reveal the obscured flows of capital and resources constituted in their construction and operation, as well as engender an understanding of data centers as tools of capitalist extraction and accumulation. I have selected these theories specifically as they operate at multiple scales, from systems-level analysis (*Technocapitalism*), to individual interactions (*Granular Geographies*).

10. Framing/Foundations - Digital Geographies

The work of Part C of this thesis is, as I said above, focused on developing a theoretical conceptual framework to account for and study data center materiality, as well as their socio-technical, political, and economic implications. While this theory is fairly narrow in scope, the project of developing this conceptualization fits squarely into the wider geographic subfield of ‘Digital Geographies’, and so it is important to situate this work within the existing scholarship. This section will review some of the theoretical approaches developed by scholars in this field, and examine the various interpretations of ‘The Digital’ within this scholarship.

While a relatively new subfield within the field of Geography, geographic research focused on the digital and digitality has been increasing rapidly (Sui & Morrill, 2004; Haefner & Strenberg, 2020). One of the most thorough reviews of the existing scholarship, and certainly a foundational paper within digital geographies is Ash, Kitchin, and Leszczynski’s (2018) ‘Digital turn, digital geographies?’, which lays out the conceptual ‘turn’ within geography, typified by the exploration of the inter-relationship between geography and the digital, and examining the geographies produced ‘through, by and of’ the digital. Before examining some of the frameworks and approaches within this field it is important to define ‘the digital’, Ash et al. (2018) offer a thorough description:

> “We use ‘the digital’, then, to make reference to material technologies characterized by binary computing architectures; the genre of sociotechno-cultural productions, artefacts, and orderings of everyday life that result from our spatial engagement with digital mediums, and the logics that both structure these ordering practices as well as their effects” (pp. 26, 2018)

Using this definition it’s clear that ‘the digital’ is at once tangible, consisting of objects, infrastructures, data, and intangible, something felt, a lived experience (McLean, 2019).

Perhaps emblematic of the need for new, robust conceptual theorizations, less than one page of Ash et al.’s (2018) paper focuses specifically on reviewing geographic scholarship related to digital infrastructures (Furlong, 2020). Their review focuses primarily on scholarship developed from the early 2000s to the early 2010's and examines the approaches scholars have taken, examining space through the concept of 'cyberspace', a realm separate from physical space (Zook et al., 2004; Hillis 1999), to the network-oriented approaches which focused on mapping and analyzing network distributions (Graham & Martin, 2001; Graham, 1998; Malecki, 2002). The scholarship here is not focused solely on the infrastructure as a distinct object and attempts to attend to the political and historical concepts of their development and operation. However, lacking in this domain is a focus on materiality and consideration of how materiality is intertwined and inseparable from political and economic considerations. Additionally, the conceptual approaches of previous work seem outdated, stemming mainly from a time before the advent of wide-scale cloud computing, and the advent of 'Big Data' (Thatcher et al., 2018).

Working within the research gaps identified above, I develop a new conceptualization to understand data centers through their material and more than material nature and impacts. The theoretical frameworks I am employing to build out this conceptualization are relevant, and critically oriented, framed by an understanding of space and materiality as intertwined with and inseparable from capitalist production, extraction, and accumulation.

11. Cross-Disciplinary Theory Synthesis

In order to expand upon the analysis I have already undertaken (Part B), and to attend to the material spatial research gaps identified in the literature review (Pickren 2018; Furlong, 2020; Sadowski, 2019, and the previous section, I think it is essential to adapt some existing theories to the study of data centers. By doing so, we can synthesize an understanding of this infrastructure which will enable us to study the flows of capital, data and resources, and the linked layered infrastructures which exist as distinct spatial nodes. To do this I will draw upon several different theories, namely: Technocapitalism (Suarez-Villa, 2009; 2016), the Planetary Mine (Arboleda, 209; 2020), and Granular Geographies (Jamieson, 2020), which encompass systemic understandings of capital and technology, entangled materialities, and multi-scalar methods to provide a holistic and powerful analytic. Below I begin by briefly describing the theories above, synthesizing the approaches, and their key interventions. I then coalesce these theories, and apply them specifically to data centers, considering how different theories and ideas add robust insight to the study of data centers, their layered impacts, (in)visibilities (Furlong, 2020), and systemic connections.

11.1 *Technocapitalism*

In the main literature review (*Section 2*) I discussed how Furlong, Hu, and others note that in order to understand data centers and other networked objects, we shouldn't focus solely on the object itself, but root the development and operation of infrastructural objects within wider systemic social and economic histories. So I will begin by grounding modern large and hyperscale data centers within modern technocapitalism, defined by Luis Suarez-Villa as, "a new form of capitalism that is heavily grounded on corporate power and its exploitation of technological creativity. (pp. 4, 2009). In his writing on technocapitalism, Suarez-Villa discusses how corporations have used and developed technology in order to seek profits from knowledge and creativity, and that:

> The values of corporatism are embedded in the research agendas and design of technology. Technological rationality is therefore not really "neutral" or "functional." Such rationality is also social, political, economic, and cultural, and it represents the

> power, the values, and interests of the dominant power: technocapitalist corporatism. (pp. 7, 2009).

This mode of capitalist production, premised on technological innovation for the purpose of corporate profits and power is highly evident in the modern data center, in particular, the cloud computing data centers, in which three companies Amazon (AWS), Google, and Microsoft control more than 60% of the total cloud computing market, which itself grew more than 32% over the course of 2020 and is now estimated to be worth more than 40 billion dollars (Canalys, 2021).

Suarez-Villa centers networks in his formulation of technocapitalism, noting that "Networks are the means through which the vital processes of technocapitalism are articulated." (pp. 56, 2009) functioning as interdependent infrastructural 'organs' (Marx, 1977). The network and network infrastructures such as data centers can be said to be carrying out the operationalization of technocapitalist shifts in social relations. The wider the extent of the network a corporation controls, the more power they have over not only the production of knowledge but its commodification. While Suarez-Villa first conceptualized technocapitalism before cloud computing grew to what it is today and before the wide-scale extraction, commodification, and accumulation of big data across space and time, we can see how what he calls the "commodification of knowledge and creativity" (2009) can be applied to the role that data centers play in the capture of personal and industrial data for the purpose of commodification, and profit. Likewise, the operation of data centers to execute AI and ML algorithms for corporations is in its own right a sort of knowledge creation and commodification, rooted in corporatism and control.

Since publishing his book on technocapitalism in 2009, other scholars have taken the concept of technocapitalism and continued to update and improve upon the ideas he presents. In particular, Erin McElroy, in their Ph.D. dissertation coined the term "racialized technocapitalism" (McElroy, 2019) as a concept that captures not only the corporate power embodied in modern technology development but also to analyze the racialized impact of the development and deployment of these technologies. They draw upon the work of Cedric Robinson's Black Marxism where he writes that capitalism and race have always been co-constituted (Robinson, 1983). Likewise Ruth Wilson Gilmore (2018), also drawing from Robinson, writes that "Capitalism [is] never not racialized". With this in mind, I think it is important to root data centers within the legacies of industrial capitalism which has led to the development of technocapitalism, of which data centers are a key component, and that technocapitalism, historically and presently is racialized in its impacts and effects. This is the first component of my theoretical conceptualization of data centers, which before we chart a framework for critically examining their impacts we must first ground them within a history of capitalist development.

11.2 ***Planetary Mine***

Now that I have established that data centers typify and embody the logic of racialized technocapitalism, I can begin to explore their operation, which I discussed in the previous section as predicated on their function as spaces of extraction and accumulation, largely in the form of data, but also material resources in the form of raw materials which make up the servers and the buildings which house data centers. To do so I draw upon the recent work of Martin Arborella (2019; 2020) who conceives of the idea of a "Planetary Mine". Building upon and theorizations of planetary-scale extraction (Labban, 2014) and conceptual logistics networks (Mezzadra and Neilson, 2017) among others, Arborella discusses how the modern mine is not bound to one fixed location, and may not even by a physical mine at all, but rather that we should "transcend the territoriality of the mine" (pp.29, 2020). Instead of focusing on one mine or location, he conceives of a planetary scale mine which is comprised of a "dense network of territorial infrastructures and spatial technologies vastly dispersed across space" (pp. 5, 2020), similar to the work of Furlong and Pickren, Arborella grounds the mine as an object within a wider system of capitalist development which shapes and drives the creation of particular modes of extraction and accumulation.

While Arborella is focused specifically on actual mines and mining logistics, I think much of what he writes can also be applied to data centers, and that it is helpful to view the data center as a sort of node within the planetary mine. I mean this literally in that AI and cloud computing has been growing significantly within the mining industry, with cloud computing being described as the "backbone of the digital mining transition" and driver of mining automation (Hiltz, 2020), but also in a conceptual sense. Data centers, which are linked by and through a global network of information communication technology infrastructures function as mines in their own right, as spaces of data mining through the extraction of personal and consumer data, and the "refining" (metaphorically) of raw data into commodified "Big Data". Jim Thatcher writes that it is "only when millions and billions of individual pieces of data are linked together algorithmically that the commodity known as big data emerges" (Thatcher et al., pp. 99, 2016), in this way data centers need to extract as much information as possible through various computational means, and at a number of different scales in order to commodify data, and accumulate not only the value-laden information that Big Data provides but the super profits which can be reaped from its transformation into capital value.

Conceiving of data centers as intrinsic with, and part of the Planetary Mine also helps to conceptualize their materiality in more finite and empirical ways, which remains a crucial task in critical data studies research. Without resource mining and extraction and many different levels, the modern data center would not exist, and as such their development and proliferation is dependent upon continued mining of many different resources and raw materials. For example, the servers which constitute the data centers computing infrastructure require a range of precious and highly finite resources including Gold, Palladium, and Cobalt, among many others (EU RMIS, 2020) that must be mined, oftentimes by exploited or even indentured labor in the Global South (Council on Foreign Relations, 2020). As cloud computing and data centers continue their exponential growth, more and more of these natural resources will be required, and so while the West, its corporate elites, and everyday citizens may economically benefit from the growth of data centers, those gains must be offset by the exploitation of cheap labor and a continuation of the imperial logics of resource extraction and flow from the global south to the global north. This uneven and unequal relationship is an example of the types of dichotomies that emerge within the system of racialized technocapitalism, which mirror and are built upon the same structures which enabled the rapid industrialization of many countries in the Global North, at the expense of communities and their ecological well being on the periphery. The critical link between the supposed 'ephemeral' cloud and raw materials cannot be overstated, and centering the mine and resource extraction, while considering the material nature of data centers can help in one way to reveal the layered (in)visibilities of this infrastructure (Furlong, 2020).

11.3 *Granular Geographies*

A final theory which I will bring into this discussion is the idea of "Granular Geographies" (Jamieson, 2020) which I think is a useful way to conceptualize data centers, as granular systems. Jamieson theorizes granular geographies as an intervention in materialist geographies and explores the topic of sand extraction through granular dynamics of "force chains, friction and phase transitions" (pp.1, 2020). In his piece, Jamieson seeks to make discrete the process of sand extraction, commodification, and larger processes such as urbanization, which requires a massive amount of sand for the purpose of construction projects and the like. He states that granular systems, "can become a critical tool for thinking with unstable materialities that seemingly gather of their own volition and crumble beneath our grasp, yet only do so under certain (*but not fully predictable*) conditions." (pp. 3, 2020) and in his article proposes how this theory could be applied to other opaque systems of flows and relations, including data.

I'll briefly review the components of granular systems as laid out by Jamieson, and then discuss how data centers function within a hypothetical system of data granularity.

Force Chains

Jamieson begins by considering the role of "force chains" within granular systems and defines this as the "interconnected and intercalated ecologies of 'production' and 'consumption' intrinsic to resources. (pp. 6, 2020) and the often highly obscured chains of commodification that transform raw material into a value-laden commodified object. He notes that " there is a necessary coherence of knowledge-production in extractive industries that serves to rationalize a complex and messy geopolitical process and minimize its externalities" (pp. 7, 2020). He applies to sand in that the process of extracting and commodifying sand is hard to track due to the bulk of the resource being extracted, the international network of transportation, global markets, and because the spaces of sand extraction are often different from the embedded spatiality of its use.

I believe that the force chains of data are similarly obscured, and also similarly constructed. Like sand, individual data points are highly numerous and it's only through the assemblage of many data points that the commodification of data can occur. Likewise, the spaces of data extraction, through personal browsing habits, user-contributed data, industrial or commercial data or Volunteered Geographic Information is usually occurring away from and separate from the data center, which has its own unique spatial and material composition and role within the force chain.

Granular Friction

Jamieson also conceives of the idea of granular friction and relates this to the idea of metabolic rift and a dialectical relationship between capital and nature in which both are co-constituted. The friction is evidenced between the ever-increasing extraction of resources (in this case sand) and the hard ecological limits of nature, whereby ever-increasing energy needs to be input into the system to ensure that the friction or dialectical relationship between the commodity and nature will be subsumed by one or the other force.

In much the same way we can see that the abundance of data being extracted, stored, and processed in data centers has grown exponentially, resulting in the need for ever larger and more powerful data centers and cloud computing facilities. This has brought data centers up against the hard limits of natural resource extraction and in particular energy use and consumption. Exploring the friction between low latency high volume data processing, with fixed ecological limits highlights the fragility of the dialectic between the cloud and its socio-ecological setting.

Phase Transitions

The last component of granular systems is that of phase transitions, which Jamieson examines through sands function as both a liquid and solid, and its interactions between the land and water, often in highly intermixed and indeterminate ways (pp. 11, 2020). He explores how sand is transformed into a territory in the form of land reclamation projects by undergoing a series of phase transitions, which are made visible through the dynamics discussed above (*force chains and friction*). Viewing sand through its transitions, "first from geophysical relationality into a resource; then into a strategic matter of national security; then into a component part of an engineering project; then into sovereign territory." (pp. 13, 2020). If we frame data in this way, it is easy to see the ubiquitous role of the data center within the phase transitions of data, the data center acts as a sort of refinery for data to transition from "raw" resource to commodity, and a space within the commodity force chain, a site of friction and embodies the dialectical relationship between capital and nature.

Conceiving of data commodification as a granular system, and situating data centers as the linchpin node within that system is, I believe, a useful theoretical framework for revealing the layered nature of this infrastructure, and understanding of data centers as spatially contingent political objects, we can begin to make legible the highly obscured capital flows which they enable, and contest the deterritorialization of the internet and platforms through which social relations are mediated, and material conditions realized.

12. Theory Formation/Adaption

"...in place of the isolated machine, a mechanical monster whose body fills whole factories, and whose demonic power, at first hidden by the slow and measured motions of its gigantic members, finally bursts forth in the fast and feverish whirl of its countess working organs."

(Marx, Capital Volume 1, Part 4, Chapter 15)

In the previous section, I focused on synthesizing three different theories; Technocapitalism, the Planetary Mine, Granular Geographies, and briefly described how these frameworks incorporate both digitality and materiality and began to explore how they operate at different scales. Here I will expand upon these theories as a whole, connecting them specifically to data center research, and explaining the fundamental dynamics and logics of data center operations that these theories can reveal and explain. In doing so I will situate this infrastructure within wider social processes, as well as discuss how this new cohesive conceptualization can be employed as an analytical framework for materialist, and critical analysis for studying data centers and other ICT infrastructures.

Just as I did with the spatial analysis in Part B, I'll begin with assessing these frameworks' function at a large scale and then narrow the scope and scale to examine how they can also be employed for individual or other small-scale analyses. I focus on what these theoretical concepts reveal about the linkages between data center development, operations, and economic processes such as globalization and corporatization. Layout how these approaches attend to the material as well as social considerations. Most importantly, I explain how these frameworks can be used for a robust analysis that considers the materiality of the data center through multiple scales, connecting the physical location of the data center to external spaces of resource extraction and accumulation, as well as data capture, processing, and commodification.

These allow us to consider the data center as a political object, composed of and determined by intractable social and economic forces. Additionally, in connecting locations of extraction and accumulation to the data centers be it resource mining for server materials, or a surveillance camera uploading to the cloud, there is also a linkage of political struggles and praxis-based interventions which inherently are connected despite the locational distance.

Fundamentally this begins with naming racialized technocapitalism as the dominant logic for capitalist production. This is the overarching system within which data centers were developed, and as such, they are deployed specifically for the purpose of capitalist production. Where data centers are located is based on decision-making around profit, through the sale of computational resources, data storage, and through the assembly of commodified data (Thatcher, 2018; Sandowski, 2019). Until the 1990s, most data centers were owned and operated by individual companies for their own data processing needs (Miniman, 2014). However, coinciding with the proliferation of post-Fordist neoliberal governance, the growth of multinational technology corporations, and the growth of "big data" the data center industry has coincided with the growth of distributed and hybrid networks, whereby companies have moved away from maintaining their own networks and data centers and relying on cloud providers (Alamouti, 2017). Now the majority of colocation data centers are owned by a handful of corporations, the cloud computing market is dominated by Amazon, Google, and Microsoft who control more than 60% of the industry, and few companies operate their own independent data centers, relying more on the major platforms for cloud computing and data storage needs (Jones, 2021).

Faced with the ever-present crisis of capital, i.e. the tendency for the rate of profit to fall, and the need for continual growth. Data centers have been operationalized to accrue profits at multiple scales both for the data center or cloud platform operator, but also the business users of those services. Data centers and cloud computing driven ‘data analytics’ are seen as a one size fits all solution, from use in agriculture to operate *smart farms* (Trilogy Networks, n.d.), to the development of *smart mines* (Howden, n.d.; Rio Tinto, n.d.), as well as for state surveillance and incarceration evidenced by contracts like the one between Clarity Partners, LLC, a cloud computing provider and the Chicago Police Department (Clarity Partners, n.d.).

Exploring the idea of ‘smart mining’ in conjunction with the concept of the Planetary Mine, it is clear that the data center can and does operate as a tool of the mining industry, disconnected from a physical mine by thousands of miles, but intrinsic to the processes of material extraction and accumulation. Operating in this way, the data center takes on a new material composition, embodying the material impacts at the site of resource extraction, while also being composed of numerous material resources in the form of microchips and server components. So it at once supports and expands the processes of the planetary mine's development, and also in a way operates as a mine itself, as a site of data capture and processing.

Finally, a fundamental aspect of this new theorization is conceiving of the data center and cloud computing broadly as exemplifying a granular system. Granular geographies allow for wide-scale or hyper-focused analysis and inherently incorporate a materialist framework and approach. Granular systems allow for researchers to conceive of the material flows made legible through or by the data center, and to examine how materials, as well as data, undergo a number of phase transitions within the data center. The data center is a territory/space where force chains coalesce, a site of friction, and instigator of phase changes as it acts on data which is captured, processed, and stored, and at the same time is composed of materials and acts on materials through similar processes.

These conceptual approaches, taken together position data centers as rooted within a system of racialized technocapitalism, dependent upon and an agent within the planetary mine, and representative of a complex granular system, whereby data and materials flow, experience friction, and undergo phase transitions. A new conceptualization linking these three approaches allows for researchers to consider the human and more than human impacts from data center operations, and situate this infrastructure more specifically within the economic and social reproductive processes of the technocapitalist reality. Not a part of a distributed and open internet system, but the privatized means of production, and ever more entrenched in the projection of power and state dominance.

This conceptualization is not to demonize computing as a whole, nor to argue that data centers or cloud computing should or should not exist, but rather that the data center as it is currently constructed, operationalized, and utilized cannot be separated from the system of capitalist production. While these technologies can and are used for humanitarian (NGO's, Philanthropy) or research purposes (this Thesis, countless other academic projects), their primary function is to deliver profits and accrue capital through extraction and accumulation at multiple scales.

13. Part C Conclusion

Through the synthesis of the theoretical framings above, we have rooted the development and operation of data centers within the dominant mode of racialized technocapitalist production, linked them as a node within the planetary mine of extraction (resources and data), and constituted them as a part of a granular system of data relations. This intervention is being made because data centers and their material and more than material impacts are an understudied topic within geography, and while it is good and useful to name this gap, it is also important to prioritize theorizing frameworks to fill said research gaps as well. These sections (*Part C*), along with the spatial and temporal analysis of data centers (*Part B*) attempt to do that, exploring not only the material impacts of data centers through their land-use patterns, but also framing data centers as political objects with long legacies and connections to systemic processes of data commodification, resource extraction, and capital accumulation.

Part D - Project Conclusion

14. Thesis Conclusion

Returning to the research questions that shaped this work: Where and in what form/s can data centers be found, and how can spatial analysis be used to locate and understand data center materiality? This research has borne out that, nationally data centers tend to be clustered in and around cities, at the city level, data centers are highly clustered around highways, near airports, and are spatially contingent with logistics and warehousing infrastructure. Using remote sensing methods to sense and study the temporal land-use and land-change patterns proved to be an effective method to examine materiality and to conduct a thorough analysis of data centers without relying on piece-mail or biased information from data center operators. These methods were effective at multiple scales, affording the opportunity to situate not only the spatial nature of data centers, but also to view the materiality of the cloud through its land use. What was missing to make this more robust was a more refined set of material analyses that would enable the study of the ecological and social implications of data centers with more specificity, as well as a coherent and useful theoretical framework.

Part C attended to this second need, namely formulating a coherent and useful theoretical and conceptual framework that would elucidate the logics of data centers within the system of capitalist production, as well as reveal more about the layered and overlapping material connections between the cloud, and global sites of extraction and accumulation. The theory synthesis and adaptation work of Part C identified that data centers are developed and oriented to serve the needs of technocapitalist production, they're linked intrinsically to the process of the planetary mine and represent a granular system wherein materials and data interact and are transformed. This formulation allows us to conceive of data centers as sites of extraction and accumulation themselves, locations where data is processed, and also tools for extraction elsewhere.

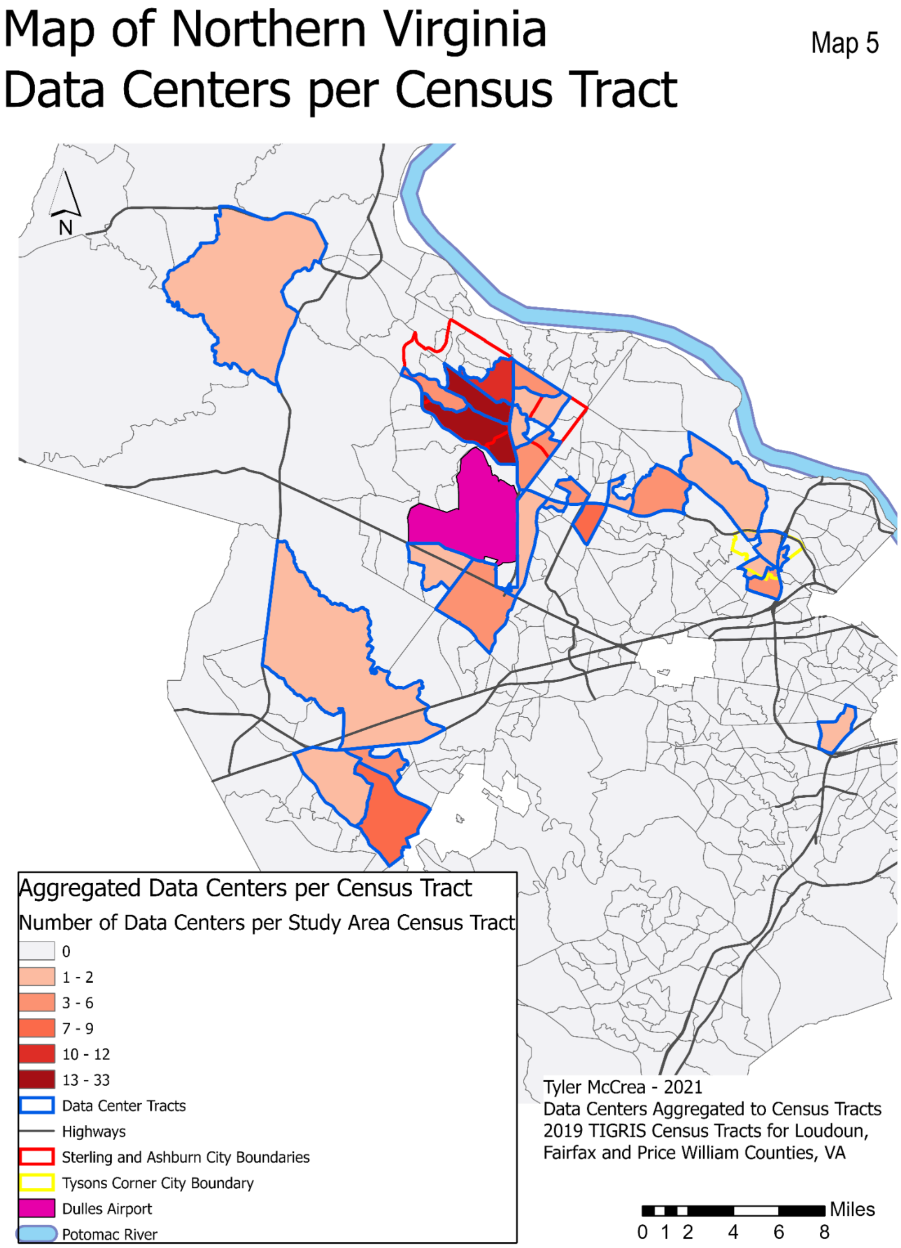
This thesis is, in many ways the first step towards a larger project that takes the successful aspects of the spatial and temporal analysis data centers, and links together those methods with the conceptual framework from Part C. By examining data centers and the cloud broadly through a granular materialist approach and analyzing the multi-scalar territories identified in Chapter 2 using remote sensing methods. The granular framework coupled with an understanding of the spaces of data center territorialization will allow for a unique analysis of the material of digital. I am imaging viewing the data center, the "mine", the device, and the community as different "study areas", coupling remote sensing with social and economic data to research both the material, and the political-economic dimensions of data centers and their role in material and data extraction, and commodification. Building off of the work done here, and going forward with the knowledge gained through this process, I think future work could be more comprehensive, as well as more rigorous in analyzing the material change and impacts.

Granular geographies in particular is a very useful materialist analytic that can allow for research into the various materials in many different forms and states which compose the data centers servers and buildings, as well as to configure data itself as a material which is transformed and acts upon the data center in particular ways. By configuring a study that explores the materiality of the physical data center, as well as the materiality and composition of data commodification, along with an understanding of the dynamics of technocapitalism and the conceptualization of the planetary mine, a future study could take this work to the next level in terms of its scope as well as outcomes.

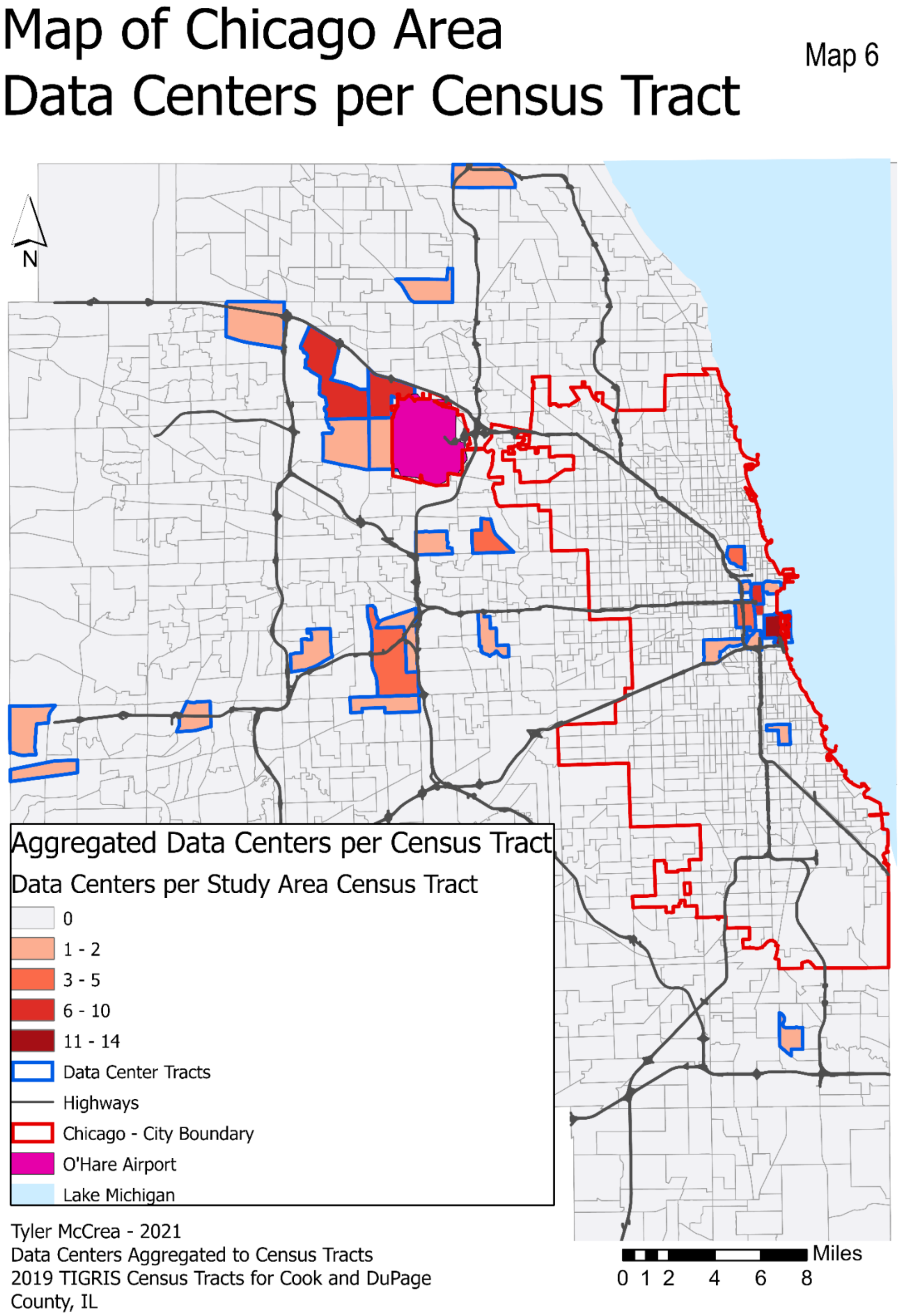
Appendices

Appendix A - Additional Maps

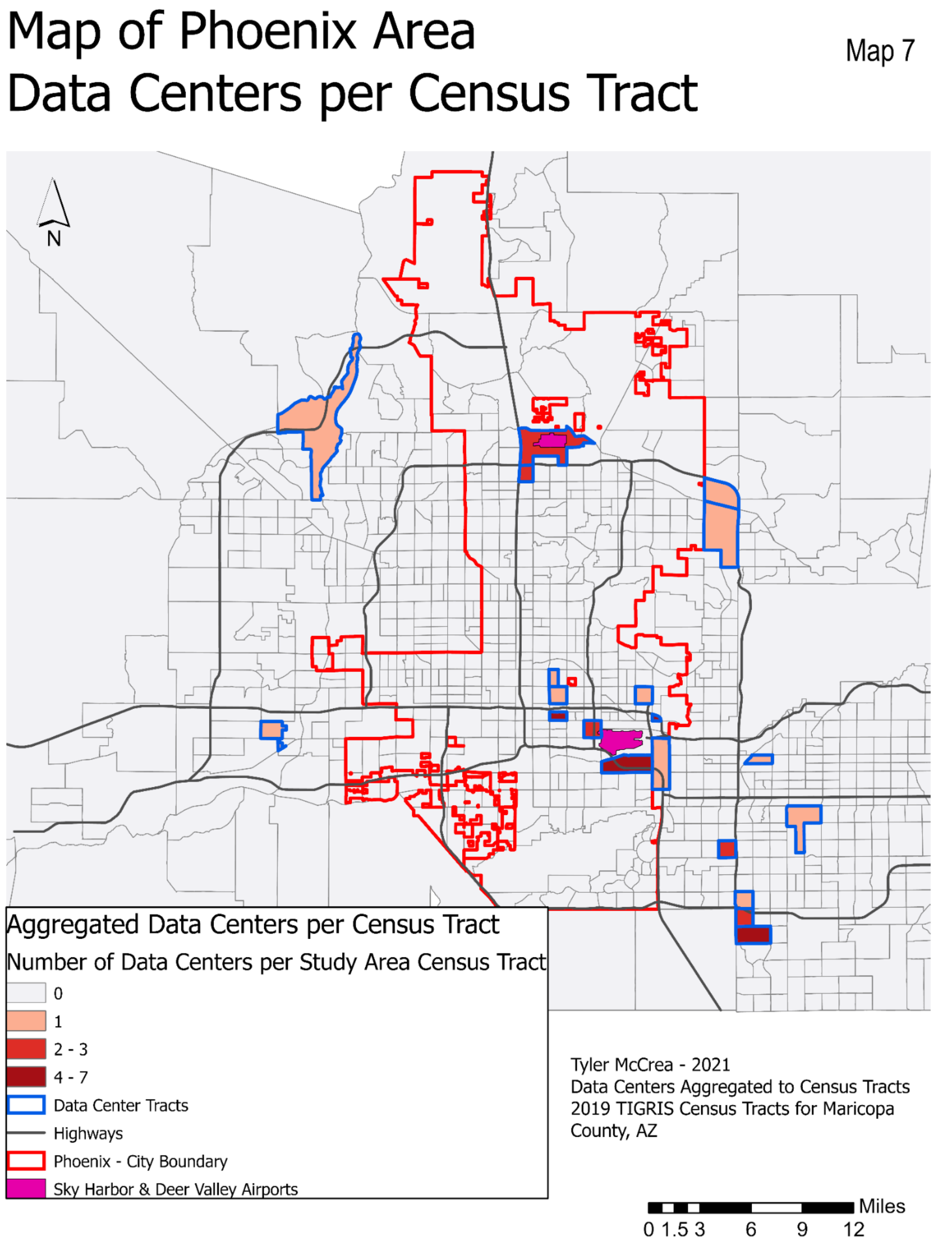
❑ Map 5 - Data Centers Aggregated to Census Tracts - Northern Virginia



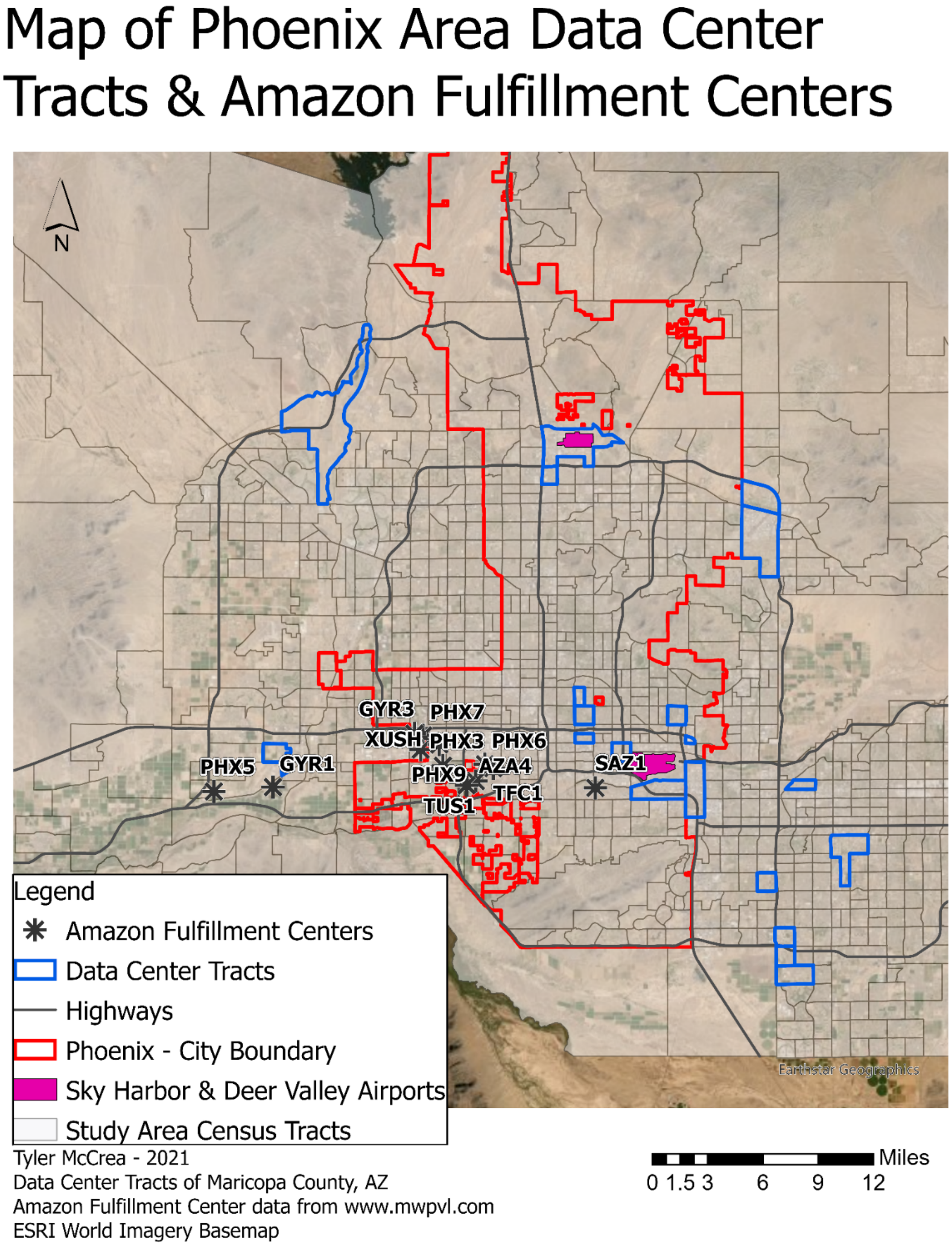
❑ Map 6 - Data Centers Aggregated to Census Tracts - Chicago Area



❑ Map 7 - Data Centers Aggregated to Census Tracts - Phoenix Area



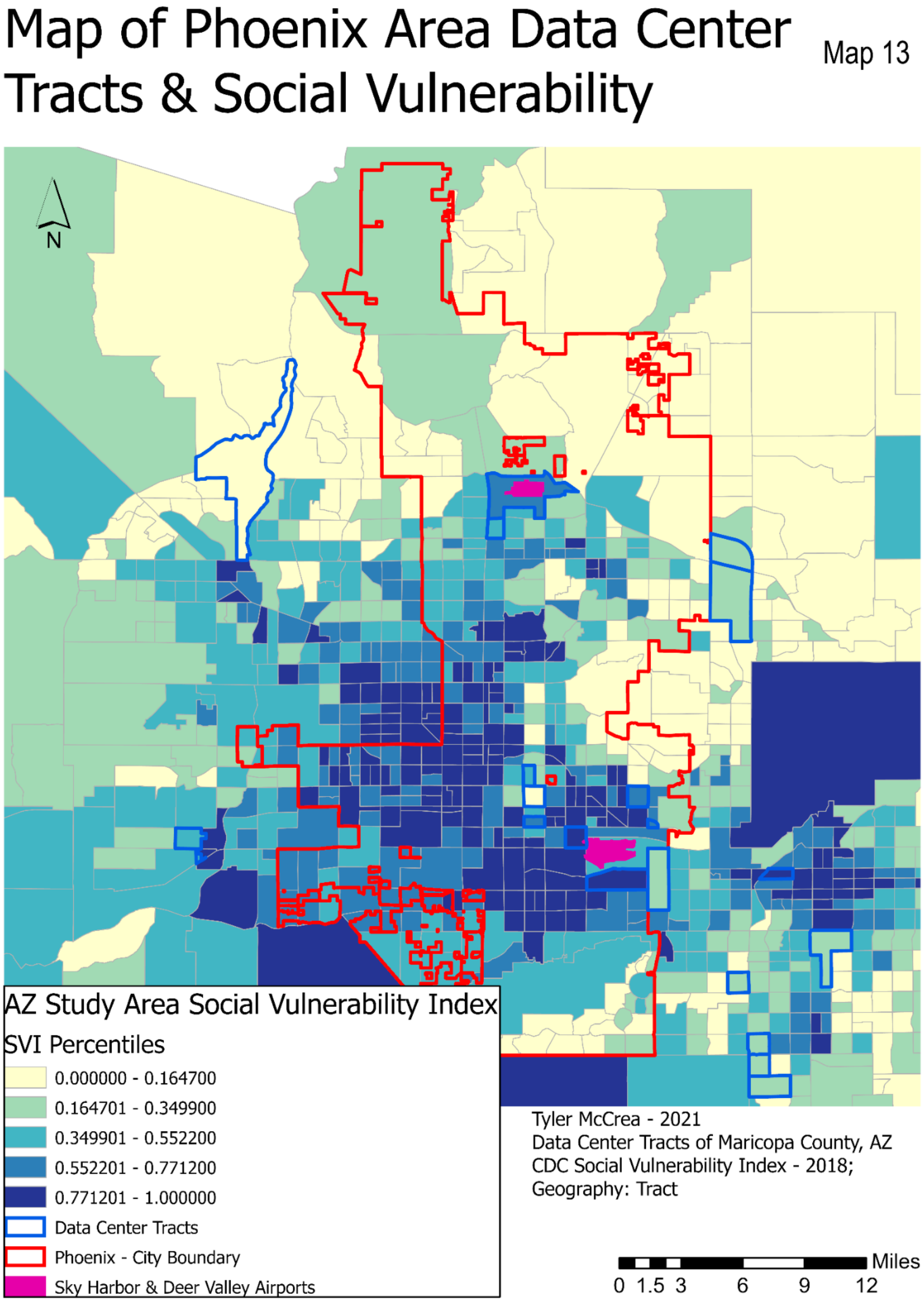
- ❑ **Map 10 - Map of Phoenix Area Data Center Tracts and Amazon Fulfillment Centers**



Tyler McCrea - 2021
Data Center Tracts of Maricopa County, AZ
Amazon Fulfillment Center data from www.mwpvl.com
ESRI World Imagery Basemap

Map 10

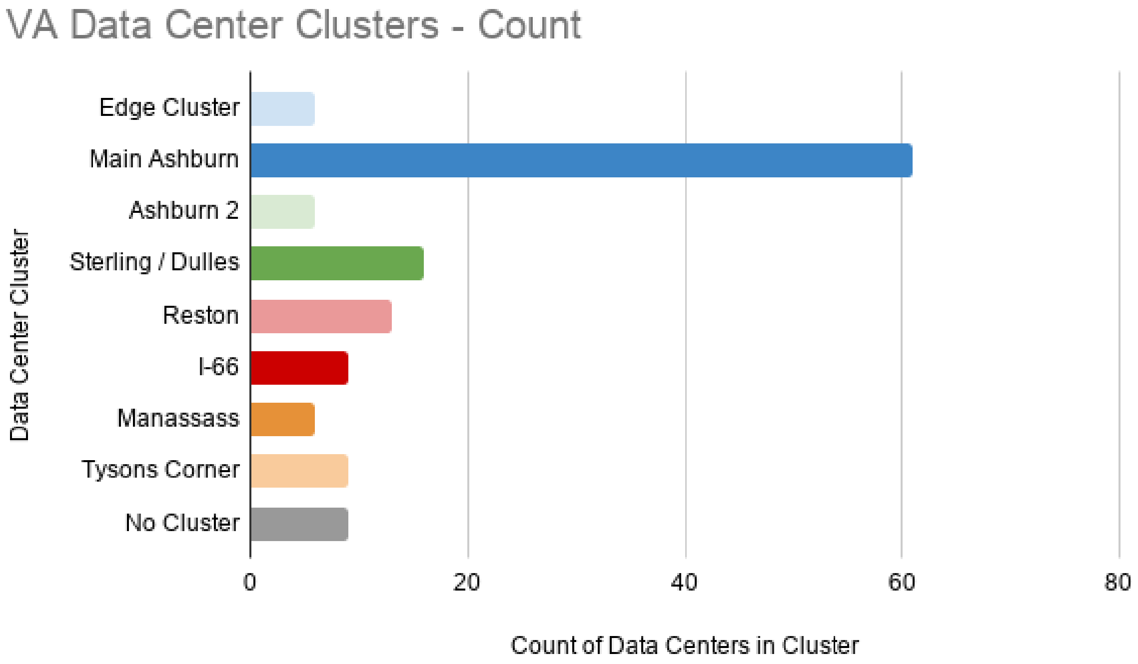
❑ Map 13 - Map of Phoenix Area Data Center Tracts & Social Vulnerability



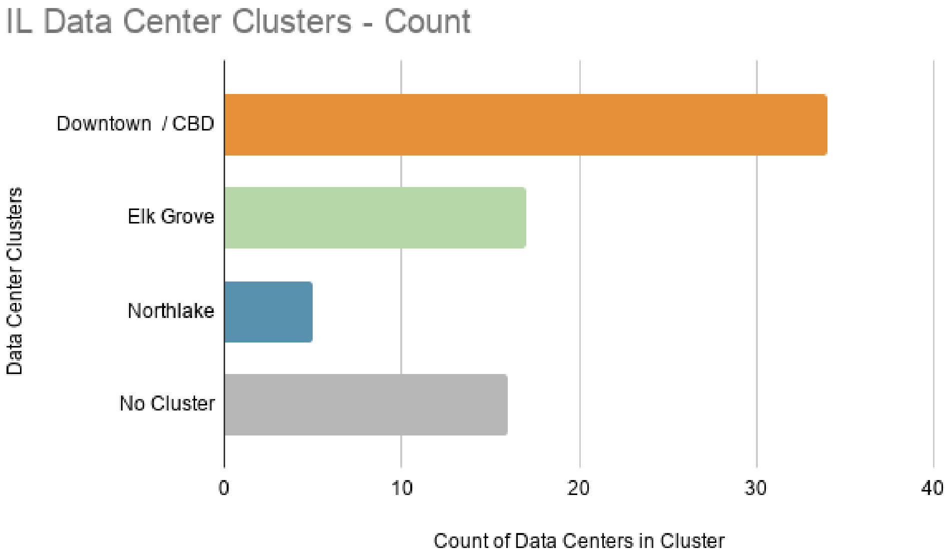
Tyler McCrea - 2021
Data Center Tracts of Maricopa County, AZ
CDC Social Vulnerability Index - 2018;
Geography: Tract

Appendix B - Figures

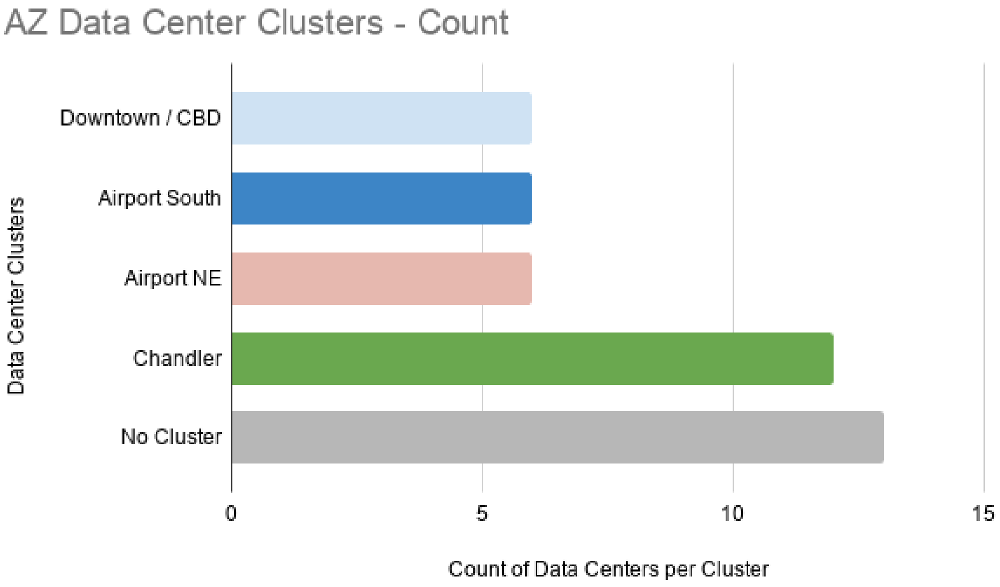
- **Figure 1 - Northern Virginia Study Area Data Centers per Cluster (Count)**



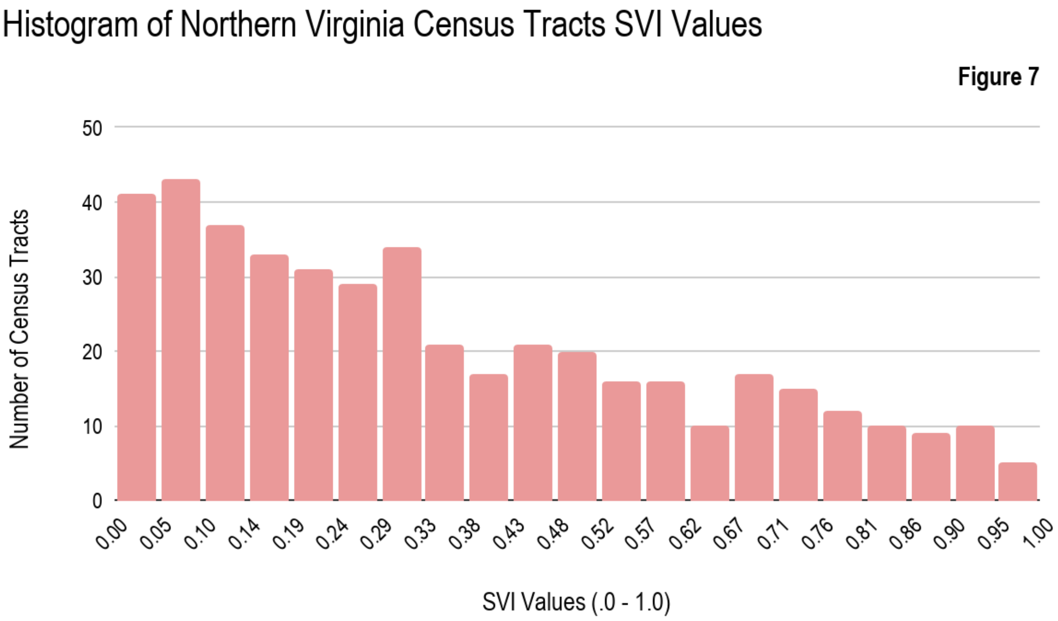
- **Figure 2 - Chicago Metro Study Area Data Centers per Cluster (Count)**



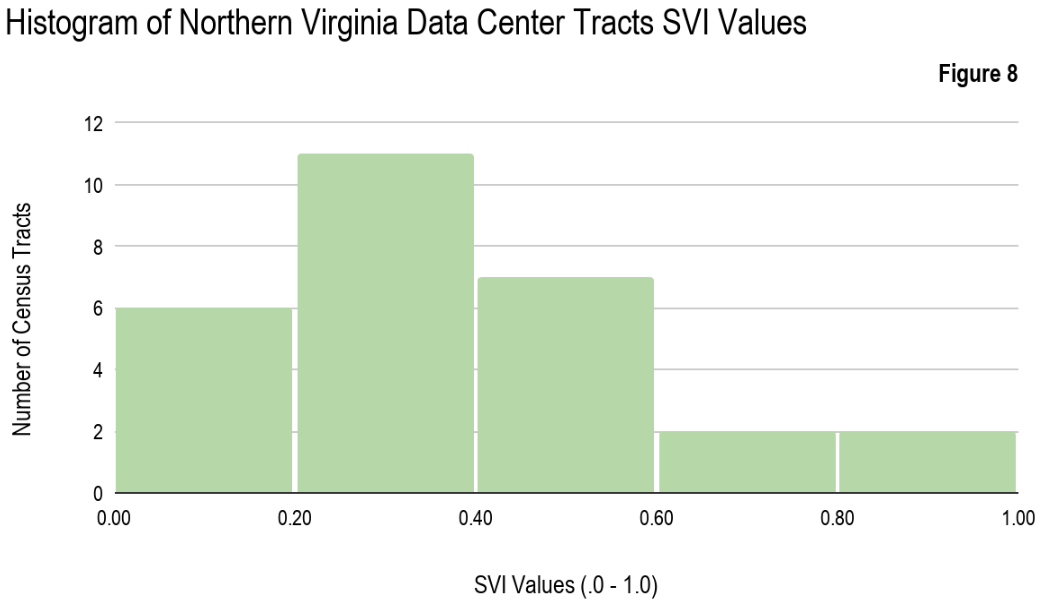
❑ **Figure 3 - Phoenix Metro Study Area Data Centers per Cluster (Count)**



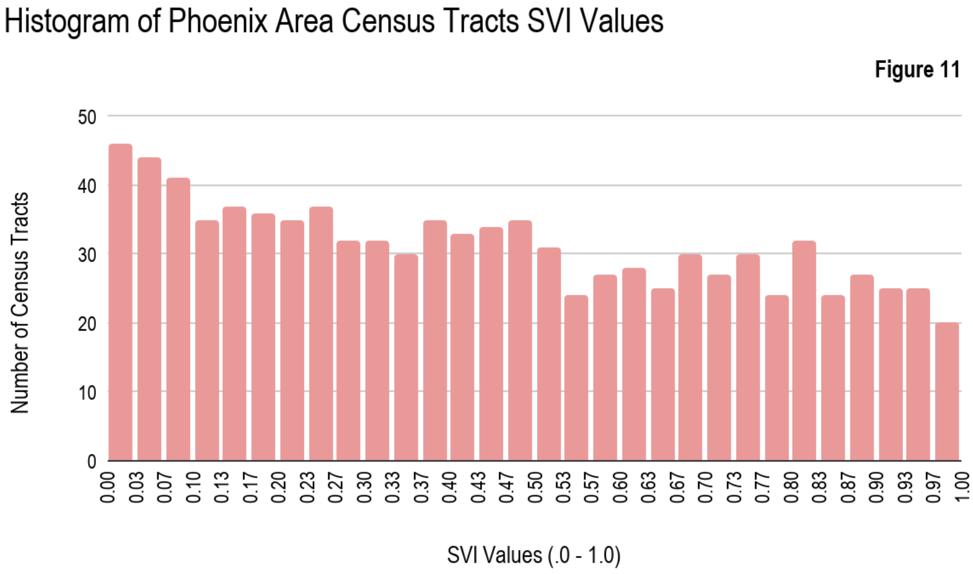
❑ **Figure 7 - Northern Virginia Study Area Histogram of All Tracts SVI Values**



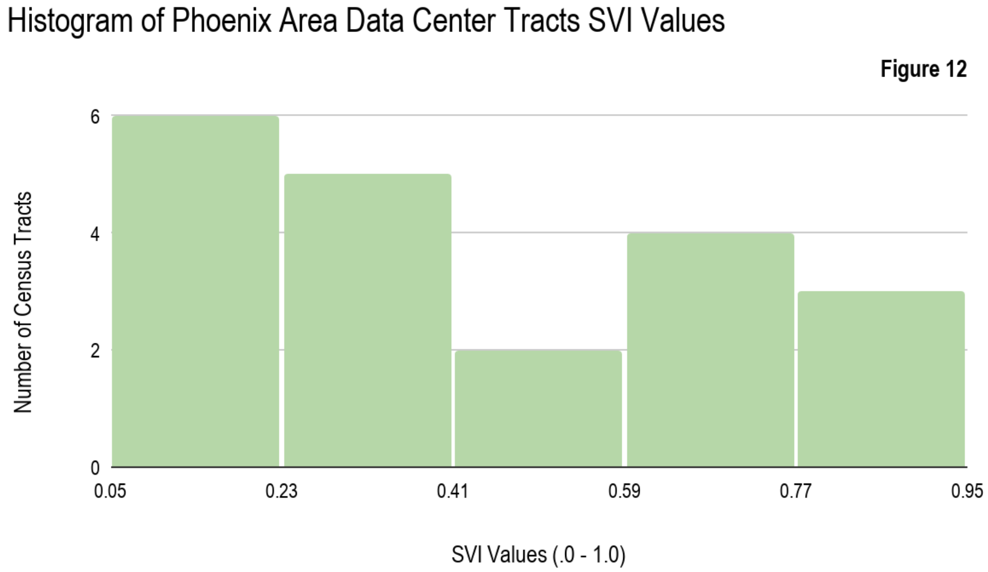
❑ **Figure 8 - Northern Virginia Study Area Histogram of Data Center Tracts SVI Values**



❑ **Figure 11 - Phoenix Metro Study Area Histogram of All Tracts SVI Values**



❑ **Figure 12 - Phoenix Metro Study Area Histogram of All Tracts SVI Values**

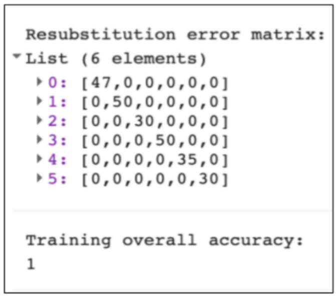


❑ **Figure 18 - Land Use Classification Error Matrices**

Landsat 8 & Sentinel-2 Classified Images Error Matrices

Landsat 8 Classification Error Matrix

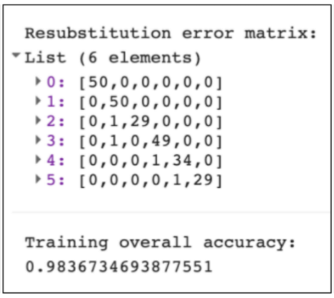
2015



```
Resubstitution error matrix:
▾List (6 elements)
  ▸0: [47,0,0,0,0,0]
  ▸1: [0,50,0,0,0,0]
  ▸2: [0,0,30,0,0,0]
  ▸3: [0,0,0,50,0,0]
  ▸4: [0,0,0,0,35,0]
  ▸5: [0,0,0,0,0,30]

Training overall accuracy:
1
```

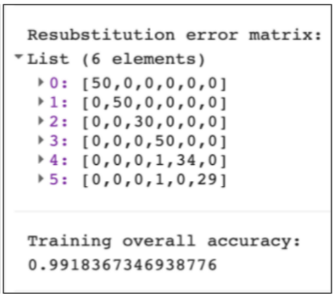
2017



```
Resubstitution error matrix:
▾List (6 elements)
  ▸0: [50,0,0,0,0,0]
  ▸1: [0,50,0,0,0,0]
  ▸2: [0,1,29,0,0,0]
  ▸3: [0,1,0,49,0,0]
  ▸4: [0,0,0,1,34,0]
  ▸5: [0,0,0,0,1,29]

Training overall accuracy:
0.9836734693877551
```

2019

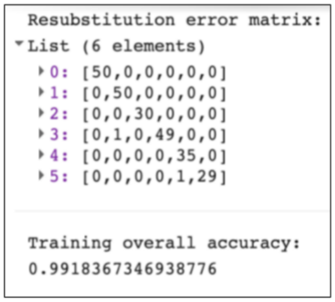


```
Resubstitution error matrix:
▾List (6 elements)
  ▸0: [50,0,0,0,0,0]
  ▸1: [0,50,0,0,0,0]
  ▸2: [0,0,30,0,0,0]
  ▸3: [0,0,0,50,0,0]
  ▸4: [0,0,0,1,34,0]
  ▸5: [0,0,0,1,0,29]

Training overall accuracy:
0.9918367346938776
```

Sentinel-2 Classification Error Matrix

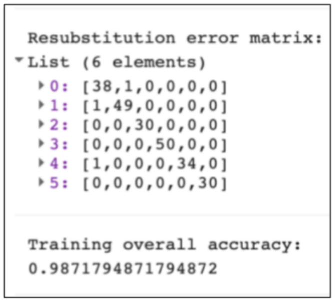
2015



```
Resubstitution error matrix:
▾List (6 elements)
  ▸0: [50,0,0,0,0,0]
  ▸1: [0,50,0,0,0,0]
  ▸2: [0,0,30,0,0,0]
  ▸3: [0,1,0,49,0,0]
  ▸4: [0,0,0,0,35,0]
  ▸5: [0,0,0,0,1,29]

Training overall accuracy:
0.9918367346938776
```

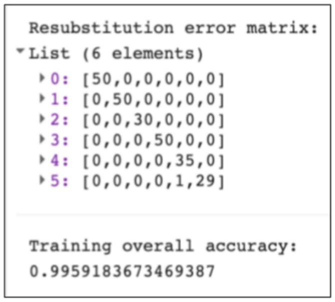
2017



```
Resubstitution error matrix:
▾List (6 elements)
  ▸0: [38,1,0,0,0,0]
  ▸1: [1,49,0,0,0,0]
  ▸2: [0,0,30,0,0,0]
  ▸3: [0,0,0,50,0,0]
  ▸4: [1,0,0,0,34,0]
  ▸5: [0,0,0,0,0,30]

Training overall accuracy:
0.9871794871794872
```

2019



```
Resubstitution error matrix:
▾List (6 elements)
  ▸0: [50,0,0,0,0,0]
  ▸1: [0,50,0,0,0,0]
  ▸2: [0,0,30,0,0,0]
  ▸3: [0,0,0,50,0,0]
  ▸4: [0,0,0,0,35,0]
  ▸5: [0,0,0,0,1,29]

Training overall accuracy:
0.9959183673469387
```

Figure 18: Contained above are the resubstitution error matrices generated by Google Earth Engine for the six classified composite images for the study area.

Appendix C - Tables

❑ Table 2 - Remote Sensing Land Use / Land Change Classification Temporal Class Changes

Data Center Class by Year Landsat 8 & Sentinel-2 Classified Composite Images

Data Centers		*Landsat 8*			*Sentinel-2*		
ID	***Name***	***2015***	***2017***	***2019***	***2015***	***2017***	***2019***
0	*44274 Round Table Plaza Data Center*	*2*	*5*	*5*	*3*	*5*	*5*
1	*44372 Round Table Plaza Data Center*	*1*	*4*	*5*	*1*	*4*	*5*
2	*Ashburn Data Center*	*3*	*3*	*3*	*3*	*3*	*3*
3	*AWS IAD71 Data Center*	*5*	*5*	*5*	*5*	*5*	*5*
4	*AWS IAD60 Ashburn Data Center*	*5*	*5*	*5*	*5*	*5*	*5*
5	*Ashburn Data Center*	*5*	*5*	*5*	*5*	*5*	*5*
6	*Ashburn Data Center*	*5*	*5*	*5*	*5*	*5*	*5*
7	*Ashburn Data Center*	*5*	*5*	*5*	*5*	*5*	*2*
8	*AWS IAD50 Ashburn Data Center*	*5*	*5*	*5*	*5*	*5*	*5*
9	*IAD3 - Ashburn I Data Center*	*5*	*5*	*5*	*5*	*5*	*5*
10	*Intergate Ashburn Data Center*	*2*	*5*	*5*	*3*	*5*	*5*
11	*XFERNET / Equinix DC3*	*5*	*5*	*5*	*5*	*5*	*5*
12	*DC3 - Ashburn Data Center*	*5*	*5*	*5*	*5*	*5*	*5*
13	*DC3 Ashburn Data Center*	*5*	*5*	*5*	*5*	*5*	*5*
14	*Northern Virginia 44521 Hastings Drive Data Center*	*5*	*5*	*5*	*5*	*5*	*5*
15	*Northern Virginia 44461 Chilum Place Data Center*	*5*	*5*	*5*	*5*	*5*	*5*
16	*Northern Virginia 44480 Hastings Drive Data Center*	*5*	*5*	*5*	*5*	*5*	*5*
17	*ACC4 44480 Hastings Drive Data Center*	*5*	*5*	*5*	*5*	*5*	*5*
18	*Northern Virginia 21625 Gresham Drive Data Center*	*5*	*5*	*5*	*5*	*5*	*5*
19	*Northern Virginia 44490 Chilum Place Data Center*	*5*	*5*	*5*	*5*	*5*	*5*
20	*21745 Sir Timothy Dr Data Center*	*1*	*5*	*5*	*1*	*5*	*5*
21	*VA3 Ashburn Data Center*	*5*	*5*	*5*	*5*	*5*	*5*
22	*VA1 Ashburn Data Center*	*5*	*5*	*5*	*5*	*5*	*5*
23	*VA2 Data Center*	*5*	*5*	*5*	*5*	*5*	*5*
24	*VA4 Virginia Data Center*	*5*	*5*	*5*	*5*	*5*	*5*
25	*44664 Guilford Dr Data Center*	*5*	*5*	*5*	*5*	*5*	*5*
26	*DC4 Ashburn Data Center*	*5*	*5*	*4*	*5*	*5*	*4*
27	*DC13 Ashburn Data Center*	*5*	*5*	*5*	*5*	*5*	*5*
28	*Wash DC (WDC1-Ashburn) Data Center*	*0*	*4*	*4*	*0*	*0*	*0*
29	*DC5 Ashburn Data Center*	*5*	*5*	*5*	*5*	*5*	*5*
30	*DC12 Ashburn Data Center*	*5*	*5*	*5*	*5*	*5*	*5*
31	*21715 Filigree Ct Data Center*	*5*	*5*	*2*	*5*	*5*	*5*
32	*DC2 Ashburn Data Center*	*5*	*5*	*2*	*5*	*5*	*5*
33	*Ashburn VA1 Virginia Data Center*	*5*	*5*	*2*	*5*	*5*	*5*
34	*IAD1 21715 Filigree Court Data Center*	*5*	*5*	*2*	*5*	*5*	*5*
35	*Equinix DC2 Ashburn Data Center*	*5*	*5*	*2*	*5*	*5*	*5*
36	*Northern Virginia 2 Data Center*	*5*	*5*	*2*	*5*	*5*	*5*

37	***Equinix DC1***	*5*	*5*	*5*	*5*	*4*	*5*
38	***DC1 Ashburn Data Center***	*5*	*5*	*5*	*5*	*4*	*5*
39	***Northern Virginia Data Center***	*5*	*5*	*5*	*5*	*4*	*5*
40	***Ashburn Data Center***	*5*	*5*	*5*	*5*	*5*	*5*
41	***Ashburn 3 Data Center***	*5*	*5*	*5*	*5*	*5*	*5*
42	***DC10 Ashburn Data Center***	*5*	*5*	*5*	*5*	*5*	*5*
43	***IAD2 Ashburn Data Center***	*5*	*5*	*5*	*5*	*5*	*5*
44	***DC6 Ashburn Data Center***	*5*	*5*	*5*	*5*	*5*	*5*
45	***DC11 Ashburn Data Center***	*5*	*5*	*5*	*5*	*5*	*5*
46	***IAD1 Washington DC Data Center***	*5*	*5*	*5*	*5*	*5*	*5*
47	***Ashburn 2 Data Center***	*5*	*5*	*5*	*5*	*5*	*5*
48	***Northern Virginia 43830 Devin Shafron Drive Data Center***	*5*	*5*	*5*	*5*	*5*	*5*
49	***Ashburn Data Center***	*0*	*0*	*3*	*0*	*0*	*0*
50	***Northern Virginia 43791 Devin Shafron Drive (Bldg D)***	*0*	*4*	*4*	*0*	*4*	*4*
51	***(Bldg D) Data Center***	*0*	*4*	*4*	*0*	*4*	*4*
52	***IAD1 Ashburn Data Center***	*0*	*4*	*4*	*0*	*4*	*4*
53	***Northern Virginia 43881 Devin Shafron Drive Data Center***	*0*	*0*	*0*	*0*	*0*	*0*
54	***Northern Virginia 44060 Digital Loudoun Plaza Data Center***	*5*	*5*	*5*	*5*	*5*	*5*
55	***Northern Virginia 44100 Digital Loudoun Plaza (Bldg J)***	*5*	*5*	*5*	*5*	*5*	*5*
56	***Northern Virginia 43780 Digital Loudoun Plaza Data Center***	*5*	*5*	*5*	*5*	*5*	*5*
57	***Northern Virginia 43940 Digital Loudoun Plaza (Bld. G)***	*5*	*5*	*5*	*5*	*5*	*5*
58	***AWS IAD67 Data Center***	*5*	*5*	*5*	*5*	*5*	*5*
59	***Northern Virginia Sterling V Data Center***	*3*	*5*	*5*	*3*	*5*	*5*
60	***AWS 45305 Sterling Data Center***	*2*	*5*	*5*	*2*	*5*	*5*
61	***AWS 45259 Sterling Data Center***	*5*	*4*	*5*	*3*	*5*	*3*
62	***AWS IAD32 Data Center***	*5*	*5*	*5*	*5*	*5*	*5*
63	***AWS IAD10 Data Center***	*5*	*5*	*5*	*5*	*5*	*5*
64	***NVA01 - Northern Virginia Data Center***	*4*	*5*	*5*	*4*	*5*	*5*
65	***Ashburn Data Center***	*1*	*4*	*5*	*1*	*5*	*5*
66	***AWS IAD76 Data Center***	*5*	*5*	*5*	*5*	*5*	*5*
67	***Dulles-The Vault Data Center***	*4*	*4*	*3*	*5*	*5*	*4*
68	***DC3 Sterling Data Center***	*5*	*5*	*5*	*5*	*5*	*5*
69	***AWS IAD69 Data Center***	*5*	*5*	*5*	*5*	*5*	*5*
70	***Northern Virginia Sterling I Data Center***	*5*	*5*	*5*	*5*	*5*	*5*
71	***21111 Ridgetop Cir Data Center***	*5*	*5*	*5*	*5*	*5*	*5*
72	***Ashburn 1 Data Center***	*5*	*5*	*5*	*5*	*5*	*5*
73	***AWS IAD15 Data Center***	*5*	*5*	*5*	*5*	*5*	*5*
74	***DC4 Sterling Data Center***	*4*	*4*	*5*	*0*	*4*	*5*
75	***AWS IAD16 Data Center***	*5*	*5*	*5*	*5*	*5*	*5*
76	***DC2 Sterling Data Center***	*5*	*5*	*5*	*5*	*5*	*5*
77	***AWS IAD12 Sterling Data Center***	*4*	*5*	*4*	*5*	*5*	*5*
78	***Sterling Data Center***	*5*	*5*	*5*	*5*	*5*	*5*
79	***Ashburn VA2 Virginia Data Center***	*5*	*5*	*5*	*5*	*5*	*5*
80	***AWS IAD57 Data Center***	*5*	*5*	*5*	*5*	*5*	*5*
81	***AWS IAD56 Data Center***	*5*	*5*	*5*	*5*	*5*	*5*
82	***AWS IAD51 Data Center***	*5*	*5*	*5*	*5*	*5*	*5*
83	***AWS IAD58 Data Center***	*5*	*5*	*5*	*5*	*5*	*5*

84	*DC7 Dulles Data Center*	*4*	*5*	*4*	*4*	*4*	*4*
85	*Northern Virginia Sterling IV Data Center*	*5*	*5*	*5*	*5*	*5*	*5*
86	*DC6 Sterling Data Center*	*5*	*5*	*5*	*5*	*5*	*5*
87	*DC5 Sterling Data Center*	*5*	*5*	*5*	*5*	*5*	*5*
88	*Herndon Data Center*	*5*	*5*	*5*	*5*	*5*	*5*
89	*Herndon 1 Data Center*	*3*	*3*	*3*	*5*	*5*	*5*
90	*Herndon 3 Data Center*	*5*	*5*	*4*	*5*	*5*	*5*
91	*Herndon Data Center*	*4*	*4*	*3*	*5*	*5*	*5*
92	*IAD4 13873 Park Center Rd Data Center*	*5*	*4*	*4*	*5*	*5*	*5*
93	*DC97 Herndon Data Center*	*5*	*5*	*4*	*5*	*5*	*5*
94	*AWS IAD61 South Riding Data Center*	*5*	*5*	*5*	*5*	*5*	*5*
96	*VA1-VA3 Campus Reston Data Center*	*5*	*5*	*5*	*5*	*5*	*5*
97	*IAD3 Washington DC Data Center*	*5*	*5*	*5*	*5*	*5*	*5*
98	*Reston Data Center*	*5*	*5*	*5*	*5*	*5*	*5*
99	*IAD3 12100 Sunrise Valley Drive Data Center*	*5*	*5*	*5*	*5*	*5*	*5*
100	*CoreSite VA1 Reston Virginia Data Center*	*5*	*5*	*5*	*5*	*5*	*5*
101	*IAD4 Washington DC Data Center*	*5*	*5*	*5*	*5*	*5*	*5*
102	*Northern Virginia 4030 Lafayette Center Drive*	*5*	*5*	*5*	*5*	*5*	*5*
103	*IAD2 11513 Sunset Hills Road Data Center*	*4*	*4*	*4*	*5*	*4*	*4*
104	*AWS IAD9 Chantilly Data Center*	*5*	*5*	*5*	*5*	*5*	*5*
105	*AWS IAD22 Chantilly Data Center*	*3*	*4*	*3*	*5*	*3*	*3*
106	*Chantilly Data Center*	*5*	*5*	*4*	*5*	*5*	*5*
107	*AWS IAD1 Chantilly Data Center*	*5*	*5*	*4*	*5*	*5*	*5*
108	*DC1 Virginia Data Center*	*3*	*3*	*2*	*5*	*5*	*5*
109	*Ashburn Data Center*	*3*	*3*	*2*	*5*	*5*	*5*
110	*Northern Virginia 1780 Business Center Drive*	*5*	*5*	*4*	*5*	*5*	*5*
111	*DC8 Vienna Data Center*	*4*	*5*	*5*	*4*	*5*	*5*
112	*1751 Pinnacle Dr Data Center*	*0*	*0*	*3*	*5*	*4*	*4*
113	*DCA2 McLean Data Center*	*4*	*4*	*4*	*5*	*5*	*5*
114	*Vienna Data Center*	*5*	*4*	*5*	*5*	*5*	*5*
116	*1764A Old Meadow Lane Data Center*	*5*	*5*	*4*	*5*	*5*	*5*
117	*McLean 1 Data Center*	*3*	*3*	*3*	*4*	*4*	*5*
118	*DC7 Vienna Data Center*	*5*	*5*	*2*	*3*	*3*	*3*
119	*VA1 - Northern Virginia Data Center*	*4*	*4*	*4*	*4*	*4*	*4*
120	*IAD6 7990 Science Applications Court*	*5*	*5*	*2*	*3*	*3*	*3*

References

Alamouti, Siavash. 2017. "The Future of Computing Is Decentralizing the Cloud." Data Center Knowledge. November 1, 2017. https://www.datacenterknowledge.com/industry-perspectives/future-computing-decentralizing-cloud.

Amoore, Louise. 2018. "Cloud Geographies: Computing, Data, Sovereignty." *Progress in Human Geography* 42 (1): 4–24. https://doi.org/10.1177/0309132516662147.

Anderson, Ben, and John Wylie. 2009. "On Geography and Materiality." *Environment and Planning A: Economy and Space* 41 (2): 318–35. https://doi.org/10.1068/a3940.

Andrae, Anders, and Peter M. Corcoran. 2013. "Emerging Trends in Electricity Consumption for Consumer ICT." Report. https://aran.library.nuigalway.ie/handle/10379/3563.

Andrae, Anders S. G., and Tomas Edler. 2015. "On Global Electricity Usage of Communication Technology: Trends to 2030." *Challenges* 6 (1): 117–57. https://doi.org/10.3390/challe6010117.

Arboleda, Martín. 2017. "Revitalizing Science and Technology Studies: A Marxian Critique of More-than-Human Geographies." *Environment and Planning D: Society and Space* 35 (2): 360–78. https://doi.org/10.1177/0263775816664099.

———. 2019. "From Spaces to *Circuits of Extraction* : Value in Process and the Mine/City Nexus." *Capitalism Nature Socialism*, August, 1–20. https://doi.org/10.1080/10455752.2019.1656758.

———. 2020. *Planetary Mine: Territories of Extraction under Late Capitalism*. New York, New York: Verso. https://www.amazon.com/Planetary-Mine-Territories-Extraction-Capitalism/dp/1788732960.

Ash, James, Rob Kitchin, and Agnieszka Leszczynski. 2018. "Digital Turn, Digital Geographies?" *Progress in Human Geography* 42 (1): 25–43. https://doi.org/10.1177/0309132516664800.

Bahari, Hazril Izan, and Siti Salbiah Mohamed Shariff. 2016. "Review on Data Center Issues and Challenges: Towards the Green Data Center." In *2016 6th IEEE International Conference on Control System, Computing and Engineering (ICCSCE)*, 129–34. https://doi.org/10.1109/ICCSCE.2016.7893558.

Barnett, Clive. 1998. "The Cultural Turn: Fashion or Progress in Human Geography?" *Antipode* 30 (4): 379–94. https://doi.org/10.1111/1467-8330.00085.

Bimber, Bruce. 1998. "The Internet and Political Mobilization: Research Note on the 1996 Election Season." *Social Science Computer Review* 16 (4): 391–401. https://doi.org/10.1177/089443939801600404.

Birant, Derya, and Alp Kut. 2007. "ST-DBSCAN: An Algorithm for Clustering Spatial–Temporal Data." *Data & Knowledge Engineering*, Intelligent Data Mining, 60 (1): 208–21. https://doi.org/10.1016/j.datak.2006.01.013.

Boyd, Danah, and Kate Crawford. 2011. "Six Provocations for Big Data." SSRN Scholarly Paper ID 1926431. Rochester, NY: Social Science Research Network. https://doi.org/10.2139/ssrn.1926431.

Brenner, Neil, David J. Madden, and David Wachsmuth. 2011. "Assemblage Urbanism and the Challenges of Critical Urban Theory." *City* 15 (2): 225–40. https://doi.org/10.1080/13604813.2011.568717.

Brodie, Patrick. 2020. "Climate Extraction and Supply Chains of Data." *Media, Culture & Society* 42 (7–8): 1095–1114. https://doi.org/10.1177/0163443720904601.

Burns, Ryan, Craig M. Dalton, and Jim E. Thatcher. 2018. "Critical Data, Critical Technology in Theory and Practice." *The Professional Geographer* 70 (1): 126–28. https://doi.org/10.1080/00330124.2017.1325749.

Burrington. 2016a. "Where Are Amazon's Data Centers? - The Atlantic." The Atlantic. January 8, 2016. https://www.theatlantic.com/technology/archive/2016/01/amazon-web-services-data-center/423147/.

Burrington, Ingrid. 2016b. *Reconnaissance*. Photography. https://nomegallery.com/exhibitions/reconnaissance/.

———. 2021. “What and Where Is The Internet.” January 1, 2021. https://observablehq.com/@lifewinning/what-and-where-is-the-internet.

Canalys. n.d. “Global Cloud Services Market Q1 2021.” Accessed June 3, 2021. https://www.canalys.com/newsroom/global-cloud-market-Q121?time=1622762618.

CBInsights. 2019. “The Future of Data Centers Industry l CB Insights.” CB Insights Research. January 24, 2019. https://www.cbinsights.com/research/future-of-data-centers/.

CBRE. 2020. “2021 U.S. Real Estate Market Outlook.” CBREUS. 2020. https://www.cbre.us/research-and-reports/2021-US-Real-Estate-Market-Outlook-Alternatives.

CDC. 2021. “CDC/ATSDR’s Social Vulnerability Index (SVI).” April 28, 2021. https://www.atsdr.cdc.gov/placeandhealth/svi/index.html.

Cho, Sunghwan, and May Yuan. 2019. “Placial Analysis of Events: A Case Study on Criminological Places.” *Cartography and Geographic Information Science* 46 (6): 547–66. https://doi.org/10.1080/15230406.2019.1578265.

CISCO. 2020. “Cisco Annual Internet Report.” Cisco. 2020. https://www.cisco.com/c/en/us/solutions/executive-perspectives/annual-internet-report/index.html.

“Clarity to Develop Next Generation of Law Enforcement Software Applications for Chicago Police.” 2013. Clarity Partners. September 1, 2013. https://www.claritypartners.com/clarity-to-develop-next-generation-of-law-enforcement-software-applications-for-chicago-police/.

Collins, Harry, and Robert Evans. 2002. “The Third Wave of Science Studies.” *Soc Stud Sci* 32 (April). https://doi.org/10.1177/0306312702032002003.

Crain, Matthew. 2018. “The Limits of Transparency: Data Brokers and Commodification.” *New Media & Society* 20 (1): 88–104. https://doi.org/10.1177/1461444816657096.

Cutcliffe, John R. 2000. “Methodological Issues in Grounded Theory.” *Journal of Advanced Nursing* 31 (6): 1476–84. https://doi.org/10.1046/j.1365-2648.2000.01430.x.

Dalton, Craig M, Linnet Taylor, and Jim Thatcher (alphabetical). 2016. “Critical Data Studies: A Dialog on Data and Space.” *Big Data & Society* 3 (1): 2053951716648346. https://doi.org/10.1177/2053951716648346.

Dalton, Craig, and Jim Thatcher. 2014. “What Does A Critical Data Studies Look Like, And Why Do We Care?” Society & Space. May 12, 2014. https://www.societyandspace.org/articles/what-does-a-critical-data-studies-look-like-and-why-do-we-care.

Data Farms. 2018. “Diagramming Data Centers.” 2018. https://www.datafarms.org/2020/07/24/diagramming-data-centers/.

datacenterlocations. n.d. “Data Center Locations: Top Cities, States, Countries and Regions.” Accessed June 3, 2021. https://www.datacenters.com/locations.

Dawn-Hiscox, Tanwen. 2020. “Virginia Offered Record Tax Breaks to Attract Data Centers in 2020.” December 24, 2020. https://www.datacenterdynamics.com/en/news/virginia-offered-record-tax-breaks-attract-data-centers-2020/.

DOI. 2018. “Revealing Our Planet Through Earth As Art.” October 10, 2018. https://www.doi.gov/blog/revealing-our-planet-through-earth-art.

Edgerton, David. 1995. “Merrit Roe Smith and Leo Marx (Eds.), Does Technology Drive History? The Dilemma of Technological Determinism. Cambridge, MA and London: MIT Press, 1994. Pp. Xv + 280. ISBN 0-262-19347-7, £31.50 (Hardback); 0-262-69167-1, £14.95 (Paperback).” *The British Journal for the History of Science* 28 (3): 370–72. https://doi.org/10.1017/S0007087400033355.

EIA. 2021. “Detailed State Data.” March 26, 2021. https://www.eia.gov/electricity/data/state/.

Ester, Martin, Hans-Peter Kriegel, Jörg Sander, and Xiaowei Xu. 1996. “A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise.” In *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, 226–31. KDD’96. Portland, Oregon: AAAI Press.

Eubanks, Virginia. 2018. *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*. St. Martin's Publishing Group.

Feenberg, Andrew. 2014. *The Philosophy Of Praxis: Marx, Lukács And The Frankfurt School*. Verso Books.

Fish, Adam, and Bradley Garrett. 2019. "Resurrection from Bunkers and Data Centers − Adam Fish and Bradley L. Garrett." *Culture Machine* 18 (April). https://culturemachine.net/vol-18-the-nature-of-data-centers/resurrection-from-bunkers/.

Forman, Peter. 2021. "Materialist Dialogues and the Granular." *Dialogues in Human Geography*, March, 20438206211004860. https://doi.org/10.1177/20438206211004859.

Furlong, Kathryn. 2020. "Geographies of Infrastructure II: Concrete, Cloud and Layered (in)Visibilities." *Progress in Human Geography*, May, 0309132520923098. https://doi.org/10.1177/0309132520923098.

geocodio. n.d. "Hassle-Free Geocoding - Geocodio." Accessed June 3, 2021. https://www.geocod.io/.

Ghiasi, Ali, and Rich Baca. 2014. "Overview of the Largest Data Centers." IEEE.

Gibbons, Michael. 1994. *The New Production of Knowledge : The Dynamics of Science and Research in Contemporary Societies*. London; Thousand Oaks, CA: SAGE Publications.

Gideon, Sigfried. 1948. *Mechanization Takes Command*. Minneapolis: Minnesota Press.

Gilmore, Ruth Wilson. 2018. "Abolition Geography and the Problem of Innocence." 2018. https://www.revistatabularasa.org/en/non-classe-es/abolition-geography-and-the-problem-of-innocence/.

Good Jobs First. 2021. "Mapping Amazon." ArcGIS StoryMaps. February 10, 2021. https://storymaps.arcgis.com/stories/adc5ff253a3643f88d39e7f3ef1a09ee.

Google. n.d. "Data Center Locations." Google Data Centers. Accessed June 2, 2021. https://www.google.com/about/datacenters/locations/.

Goto, Kunio. 2013. “STS and Marxist Study: Where Are We Standing Now?” *Social Epistemology* 27 (2): 125–29. https://doi.org/10.1080/02691728.2013.793755.

Graham, Stephen. 1998. “The End of Geography or the Explosion of Place? Conceptualizing Space, Place and Information Technology.” *Progress in Human Geography* 22 (2): 165–85. https://doi.org/10.1191/030913298671334137.

Graham, Steve, and Simon Marvin. 2001. *Splintering Urbanism: Networked Infrastructures, Technological Mobilities and the Urban Condition*. London: Routledge. https://doi.org/10.4324/9780203452202.

Haefner, Lukas, and Rolf Sternberg. 2020. “Spatial Implications of Digitization: State of the Field and Research Agenda.” *Geography Compass* 14 (12): e12544. https://doi.org/10.1111/gec3.12544.

Haraway, Donna J. 1997. “The Virtual Speculum in the New World Order.” *Feminist Review* 55 (1): 22–72. https://doi.org/10.1057/fr.1997.3.

Haraway, Donna Jeanne. 1991. *Simians, Cyborgs, and Women: The Reinvention of Nature*. Free Association Books.

Harding, Sandra G. 2004. *The Feminist Standpoint Theory Reader: Intellectual and Political Controversies*. Psychology Press.

Hardt, Michael, and Antonio Negri. 2005. *Multitude: War and Democracy in the Age of Empire*. Penguin.

Hillis, Ken. 2001. “Digital Sensations: Space, Identity, and Embodiment in Virtual Reality.” *University of Minnesota Press, Minneapolis, Minnesota*, January.

Hiltz, Robert. 2020. “Backbone of the Digital Mine.” April 2, 2020. https://www.miningmagazine.com/innovation/news/1384175/backbone-of-the-digital-mine.

Hogan, Mél. 2018. “Data Is Airborne; Data Is Inborn: The Labor of the Body in Technoecologies.” *First Monday* 23 (3). http://dx.doi.org/10.5210/fm.v23i3.8285.

Hogan, Mél, and Alix Johnson. 2017. “Introducing Location and Dislocation: Global Geographies of Digital Data.” *Imaginations* (blog). September 5, 2017. http://imaginations.glendon.yorku.ca/?p=9950.

Holt, Jennifer, and Patrick Vonderau. 2015. *“Where the Internet Lives”: Data Centers as Cloud Infrastructure*. *Signal Traffic*. Illinois: University of Illinois Press. https://illinois.universitypressscholarship.com/view/10.5406/illinois/9780252039362.001.0001/upso-9780252039362-chapter-3.

Howden. n.d. “The Future of Mining: Smart Mines and The IIoT.” Accessed June 2, 2021. https://www.howden.com/en-gb/articles/future-mining-smart-mines-iiot.

IDC. n.d. “Cloud to Edge Datacenter Trends.” IDC: The Premier Global Market Intelligence Company. Accessed June 2, 2021. https://www.idc.com/getdoc.jsp?containerId=IDC_P13027.

Iliadis, Andrew, and Federica Russo. 2016. “Critical Data Studies: An Introduction.” *Big Data & Society* 3 (2): 205395171667423. https://doi.org/10.1177/2053951716674238.

Jasanoff, Sheila, and Kim Sang-Hyun. 2015. *Dreamscapes of Modernity : Sociotechnical Imaginaries and the Fabrication of Power*. Chicago; London: University of Chicago Press.

JLL. 2018. “Data Center Outlook.” https://assets.recenter.tamu.edu/Documents/MktResearch/US_Industrial_DataCenterOutlook_JLL.pdf.

Johnson, Alix. 2019. “Data Centers as Infrastructural In-Betweens:” *American Ethnologist* 46 (1): 75–88. https://doi.org/10.1111/amet.12735.

Jones, Edward. 2019. “Cloud Market Share – a Look at the Cloud Ecosystem in 2021.” Kinsta. November 11, 2019. https://kinsta.com/blog/cloud-market-share/.

Kinsley, Samuel. 2014. “The Matter of ‘Virtual’ Geographies.” *Progress in Human Geography* 38 (3): 364–84. https://doi.org/10.1177/0309132513506270.

Kirsch, Scott. 2013. “Cultural Geography I: Materialist Turns.” *Progress in Human Geography* 37 (3): 433–41. https://doi.org/10.1177/0309132512459479.

———. 2020. “Historical–Geographical Materialism.” In *International Encyclopedia of Human Geography (Second Edition)*, edited by Audrey Kobayashi, 31–36. Oxford: Elsevier. https://doi.org/10.1016/B978-0-08-102295-5.10470-6.

Kitchin, Rob. 2014. *The Data Revolution: Big Data, Open Data, Data Infrastructures & Their Consequences*. Los Angeles, California: SAGE Publications.

Kitchin, Rob, and Martin Dodge. 2011. *Code/Space: Software and Everyday Life*. MIT Press.

Klein, Hans K., and Daniel Lee Kleinman. 2002. “The Social Construction of Technology: Structural Considerations.” *Science, Technology, & Human Values* 27 (1): 28–52.

Klemick, Heather, Elizabeth Kopits, and Ann Wolverton. 2017. “Data Center Energy Efficiency Investments: Qualitative Evidence from Focus Groups and Interview.” Working Paper. https://www.epa.gov/sites/production/files/2017-11/documents/2017-06_0.pdf.

Kleyman. 2015. “Logical Abstraction of the Physical Data Center.” Data Center Knowledge. February 24, 2015. https://www.datacenterknowledge.com/archives/2015/02/24/logical-abstraction-of-the-physical-data-center.

Knorr-Cetina, K. 1999. *Epistemic Cultures: How the Sciences Make Knowledge*. Cambridge, MA: Harvard University Press.

Kuhn, Thomas. 1962. *The Structure of Scientific Revolutions*. Chicago, IL: University of Chicago Press.

Kwan, Mei Po. 2016. “Algorithmic Geographies: Big Data, Algorithmic Uncertainty, and the Production of Geographic Knowledge.” *Annals of the American Association of Geographers* 106 (2): 274–82. https://doi.org/10.1080/00045608.2015.1117937.

Labban, Mazen. 2014. “Against Value: Accumulation in the Oil Industry and the Biopolitics of Labour Under Finance.” *Antipode* 46 (2): 477–96. https://doi.org/10.1111/anti.12062.

Lally, Nick, Kelly Kay, and Jim Thatcher. 2019. “Computational Parasites and Hydropower: A Political Ecology of Bitcoin Mining on the Columbia River.” *Environment and*

Planning E: Nature and Space, August, 251484861986760. https://doi.org/10.1177/2514848619867608.

Laparelli, Ippolito Pestellini. 2020. "Data Architectures." E-Flux. 2020. https://www.e-flux.com/architecture/intelligence/310404/data-architectures/.

Latour, Bruno. 1999a. "On Recalling Ant - Bruno Latour, 1999." https://journals-sagepub-com.offcampus.lib.washington.edu/doi/10.1111/j.1467-954X.1999.tb03480.x.

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

———. n.d. "Is Re-Modernization Occurring - And If So, How to Prove It?: A Commentary on Ulrich Beck." *Theory, Culture & Society* 20 (2): 35–48. https://doi.org/10.1177/0263276403020002002.

Latour, Bruno, and Catherine Porter. 1993. *We Have Never Been Modern*. Cambridge, MA: Harvard University Press. https://www.hup.harvard.edu/catalog.php?isbn=9780674948396.

Law, John. 2008. "Actor-Network Theory and Material Semiotics." In *The New Blackwell Companion to Social Theory, 3rd Edition*, edited by Bryan S. Turner, 141–58. Oxford: Blackwell. http://bookshop.blackwell.co.uk/jsp/id/The_New_Blackwell_Companion_to_Social_Theory/9781405169004.

Levenda, Anthony M, and Mahmoudi Dillon. 2019. "Silicon Forest and Server Farms: The (Urban) Nature of Digital Capitalism in the Pacific Northwest − Anthony M. Levenda and Dillon Mahmoudi." *Culture Machine*, April. https://culturemachine.net/vol-18-the-nature-of-data-centers/silicon-forest-and-server-farms/.

Li, Yuling, Xiaoying Wang, Peicong Luo, and Qingyi Pan. 2019. "Thermal-Aware Hybrid Workload Management in a Green Datacenter towards Renewable Energy Utilization." *Energies* 12 (8): 1494. https://doi.org/10.3390/en12081494.

Likas, Aristidis, Nikos Vlassis, and Jakob J. Verbeek. 2003. “The Global K-Means Clustering Algorithm.” *Pattern Recognition*, Biometrics, 36 (2): 451–61. https://doi.org/10.1016/S0031-3203(02)00060-2.

LLC, Dotsquare. n.d. “Hassle-Free Geocoding.” Geocodio. Accessed June 3, 2021. https://www.geocod.io.

Malecki, Edward J. 2002. “The Economic Geography of the Internet’s Infrastructure*.” *Economic Geography* 78 (4): 399–424. https://doi.org/10.1111/j.1944-8287.2002.tb00193.x.

Mann, Vijay, Avinash Kumar, Partha Dutta, and Shivkumar Kalyanaraman. 2011. “VMFlow: Leveraging VM Mobility to Reduce Network Power Costs in Data Centers.” In *NETWORKING 2011*, edited by Jordi Domingo-Pascual, Pietro Manzoni, Sergio Palazzo, Ana Pont, and Caterina Scoglio, 198–211. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-642-20757-0_16.

Marx, Karl. 1977. *Capital: A Critique of Political Economy, Volume One*. Vol. 1. New York, New York: Vintage Books.

Mattern, Shannon. 2018. “Maintenance and Care.” *Places Journal*, November. https://doi.org/10.22269/181120.

Mayer, Vicki. 2019. “The Second Coming: Google and Internet Infrastructure.” *Culture Machine* 18 (April). https://culturemachine.net/vol-18-the-nature-of-data-centers/the-second-coming/.

McCrea, Tyler. 2020. “Tpmccrea/Crawler_Repo: Repository for Python Webcrawler and Twitter Crawler.” 2020. https://github.com/tpmccrea/Crawler_Repo.

McElroy, Erin Mariel Brownstein. 2019. “Unbecoming Silicon Valley: Techno Imaginaries and Materialities in Postsocialist Romania.” UC Santa Cruz. https://escholarship.org/uc/item/0vt9c4bq.

McLean, J. 2020. “Feeling the Digital Anthropocene.” *Changing Digital Geographies*, 159–75. https://doi-org.offcampus.lib.washington.edu/10.1007/978-3-030-28307-0_8.

Mezzadra, Sandro, and Brett Neilson. 2017. “On the Multiple Frontiers of Extraction: Excavating Contemporary Capitalism.” *Cultural Studies* 31 (2–3): 185–204. https://doi.org/10.1080/09502386.2017.1303425.

Miniman, Stuart. 2014. “The Data Center: Past, Present And Future.” The Data Center: Past, Present and Future. February 6, 2014. http://wikibon.org/wiki/v/The_Data_Center:_Past,_Present_and_Future.

Morsello, Barbara. 2016. “Digitalization of Life — How Technology Redefine the Self in the Global Context.” *EASST Review* 35 (4). https://easst.net/article/digitalization-of-life-how-technology-redefine-the-self-in-the-global-context/.

Müller, Martin. 2015. “Assemblages and Actor-Networks: Rethinking Socio-Material Power, Politics and Space.” *Geography Compass* 9 (1): 27–41. https://doi.org/10.1111/gec3.12192.

MWPVL International. n.d. “Amazon Distribution Network Strategy.” Accessed June 3, 2021. https://mwpvl.com/html/amazon_com.html.

NASA. n.d. “What Is Remote Sensing? | Earthdata.” Accessed June 3, 2021. https://earthdata.nasa.gov/learn/backgrounders/remote-sensing/.

Naschert, Camilia, and Samantha Tomasczcyk. n.d. “Big Tech in Front as Green Power Becomes ‘must-Have’ for European Data Centers.” Accessed June 3, 2021. https://www.spglobal.com/marketintelligence/en/news-insights/latest-news-headlines/big-tech-in-front-as-green-power-becomes-must-have-for-european-data-centers-61162590.

Pickering, Andrew. 2010. *Science as Practice and Culture*. Chicago, IL: University of Chicago Press.

Pickren, Graham. 2018. “‘The Global Assemblage of Digital Flow’: Critical Data Studies and the Infrastructures of Computing.” *Progress in Human Geography* 42 (2): 225–43. https://doi.org/10.1177/0309132516673241.

———. n.d. “The Factories of the Past Are Turning into the Data Centers of the Future.” The Conversation. Accessed June 3, 2021.

http://theconversation.com/the-factories-of-the-past-are-turning-into-the-data-centers-of-the-future-70033.

Pinch, Trevor. n.d. "The Social Construction of Technology: A Review. In R. Fox (Ed.)." *Technological Change*, 17–35.

Pradhan, Rudra P., Girijasankar Mallik, and Tapan P. Bagchi. 2018. "Information Communication Technology (ICT) Infrastructure and Economic Growth: A Causality Evinced by Cross-Country Panel Data." *IIMB Management Review* 30 (1): 91–103. https://doi.org/10.1016/j.iimb.2018.01.001.

"Raw Materials Information System." n.d. Accessed June 3, 2021. https://rmis.jrc.ec.europa.eu/?page=crm-list-2020-e294f6.

Reinsel, David, John Gantz, and John Rydning. 2018. "The Digitalization of the World from the Edge to the Core." IDC.

RioTinto. n.d. "Smart Mining." Accessed June 2, 2021. https://www.riotinto.com/about/innovation/smart-mining.

Robinson, Cedric J. 1983. *Black Marxism: The Making of the Black Radical Tradition*. Zed.

Rodgers, Dennis, and Bruce O'Neill. 2012. "Infrastructural Violence: Introduction to the Special Issue." *Ethnography* 13 (4): 401–12. https://doi.org/10.1177/1466138111435738.

Rogers, Alisdair, Noel Castree, and Rob Kitchin. 2013. "Materiality." In *A Dictionary of Human Geography*. Oxford University Press. http://www.oxfordreference.com/view/10.1093/acref/9780199599868.001.0001/acref-9780199599868-e-1148.

Sadowski, Jathan. 2019. "When Data Is Capital: Datafication, Accumulation, and Extraction." *Big Data & Society*, January. https://doi.org/10.1177/2053951718820549.

Safransky, Sarah. 2020. "Geographies of Algorithmic Violence: Redlining the Smart City." *International Journal of Urban and Regional Research*. https://onlinelibrary-wiley-com.offcampus.lib.washington.edu/doi/full/10.1111/1468-2427.12833.

Salazar-Acosta, Monica, and Adam Holbrook. 2008. "Some Notes on Theories of Technology, Society and Innovation Systems for Science and Technology Policy Studies." Vancouver, BC: Simon Fraser University.

Sayes, Edwin. 2017. "Marx and the Critique of Actor-Network Theory: Mediation, Translation, and Explanation." *Distinktion: Journal of Social Theory* 18 (3): 294–313. https://doi.org/10.1080/1600910X.2017.1390481.

Shapin, Steven. 1996. *The Scientific Revolution*. Chicago: University of Chicago Press. https://press.uchicago.edu/ucp/books/book/chicago/S/bo29143181.html.

Shaw, Jonathan. 2019. "Technology, Paternity, Patriarchy." Harvard Magazine. February 11, 2019. https://www.harvardmagazine.com/2019/03/marriage-technology-feminism.

Sheffi, Yossi, Maria Jesus Saenz, Liliana Rivera, and David Gligor. 2019. "New Forms of Partnership: The Role of Logistics Clusters in Facilitating Horizontal Collaboration Mechanisms." *European Planning Studies* 27 (5): 905–31. https://doi.org/10.1080/09654313.2019.1575797.

Shehabi, Arman, Eric Masanet, Hillary Price, Arpad Horvath, and William W. Nazaroff. 2011. "Data Center Design and Location: Consequences for Electricity Use and Greenhouse-Gas Emissions." *Building and Environment* 46 (5): 990–98. https://doi.org/10.1016/j.buildenv.2010.10.023.

Shehabi, Arman, Sarah J. Smith, Eric Masanet, and Jonathan Koomey. 2018. "Data Center Growth in the United States: Decoupling the Demand for Services from Electricity Use." *Environmental Research Letters* 13 (12): 124030. https://doi.org/10.1088/1748-9326/aaec9c.

Soja, Edward W., and Costis Hadjimichalis. 1985. "Between Geographical Materialism and Spatial Fetishism:" *Antipode* 17 (2–3): 59–67. https://doi.org/10.1111/j.1467-8330.1985.tb00334.x.

Steffens, Will, and Timothy Comerford. 2018. "The Surge of Hyperscale Data Centers." Area Development. January 17, 2018. https://www.areadevelopment.com/data-centers/Data-Centers-Q1-2018/the-surge-of-hyperscale-data-centers.shtml.

St.Germain, Joel. 2019. “Why Is Ashburn the Data Center Capital of the World?” August 29, 2019. https://www.datacenters.com/news/why-is-ashburn-the-data-center-capital-of-the-world.

Suarez-Villa, Luis. 2009. *Technocapitalism: A Critical Perspective on Technological Innovation and Corporatism*. Temple University Press. www.jstor.org/stable/j.ctt14btdwx.

———. 2016. *Globalization and Technocapitalism : The Political Economy of Corporate Power and Technological Domination*. Routledge. https://doi.org/10.4324/9781315585123.

Sui, Daniel, and Richard Morrill. 2004. “Computers and Geography: From Automated Geography to Digital Earth.” In *Geography and Technology*, edited by Stanley D. Brunn, Susan L. Cutter, and J. W. Harrington, 81–108. Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-1-4020-2353-8_5.

Swyngedouw, Erik, Erik Swyngedouw, and Nikolas C. Heynen. 2003. *Urban Political Ecology,*.

Taylor, A.R.E. 2017. “The Technoaesthetics of Data Centre ‘White Space.’” *Imaginations* (blog). September 6, 2017. http://imaginations.glendon.yorku.ca/?p=9958.

———. 2019. “The Data Center as Technological Wilderness − A.R.E. Taylor.” *Culture Machine* 18 (April). https://culturemachine.net/vol-18-the-nature-of-data-centers/data-center-as-techno-wilderness/.

Thellwall, Mike, and David Stuart. 2006. “Web Crawling Ethics Revisited: Cost, Privacy, and Denial of Service.” *Information Technology and Society*. https://onlinelibrary-wiley-com.offcampus.lib.washington.edu/doi/full/10.1002/asi.20388.

Thinking Big Data in Geography: New Regimes, New Research. 2018. University of Nebraska Press. https://doi.org/10.2307/j.ctt21h4z6m.

Thornes, John E. 2004. “The Visual Turn and Geography (Response to Rose 2003 Intervention).” *Antipode* 36 (5): 787–94. https://doi.org/10.1111/j.1467-8330.2004.00452.x.

Tolia-Kelly, Divya. 2012. “The Geographies of Cultural Geography II: Visual Culture.” *Progress in Human Geography* 36 (February): 135–42. https://doi.org/10.1177/0309132510393318.

Trilogy Networks. n.d. “Trilogy Networks.” Trilogy Networks. Accessed June 2, 2021. https://trilogynet.com/.

Tung-Hui Hu. 2015. *A Prehistory of the Cloud*. 1st ed. The MIT Press.

US Census Bureau, US Census. n.d. “TIGER/Line Shapefiles.” The United States Census Bureau. Accessed June 3, 2021. https://www.census.gov/geographies/mapping-files/time-series/geo/tiger-line-file.html.

Varnelis, Kazys. n.d. “Harvard Design Magazine: Eyes That Do Not See: Tracking the Self in the Age of the Data Center.” Accessed June 3, 2021. http://www.harvarddesignmagazine.org/issues/38/eyes-that-do-not-see-tracking-the-self-in-the-age-of-the-data-center.

Velkova, Julia. 2019. “Data Centers as Impermanent Infrastructures.” *Culture Machine* 18 (April). https://culturemachine.net/vol-18-the-nature-of-data-centers/data-centers-as-impermanent/.

Whatmore, Sarah. 2006. “Materialist Returns: Practising Cultural Geography in and for a More-than-Human World.” *Cultural Geographies* 13 (4): 600–609. https://doi.org/10.1191/1474474006cgj377oa.

“Why Cobalt Mining in the DRC Needs Urgent Attention.” n.d. Accessed June 3, 2021. https://www.cfr.org/blog/why-cobalt-mining-drc-needs-urgent-attention.

Winner, Langdon. 1980. “Do Artifacts Have Politics?” *Daedalus* 109 (1): 121–36.

Woolgar, Steve. 1991. “The Turn to Technology in Social Studies of Science.” *Science, Technology, & Human Values* 16 (1): 20–50. https://doi.org/10.1177/016224399101600102.

Zhang, Shaozeng, Bo Zhao, Yuanyuan Tian, and Shenliang Chen. 2021. “Stand with #StandingRock: Envisioning an Epistemological Shift in Understanding Geospatial Big Data in the ‘Post-Truth’ Era.” *Annals of the American Association of Geographers* 111 (4): 1025–45. https://doi.org/10.1080/24694452.2020.1782166.

“Zoning | Loudoun County, VA - Official Website.” n.d. Accessed June 3, 2021. https://www.loudoun.gov/zoning.

Zook, Matthew, Martin Dodge, and Anthony Townsend. 2004. *New Digital Geographies: Information,.*