Indigenous Data Discourse: Categorization and Open Data in Colombia

Maria Alejandra Nino Aguirre
New Media and Digital Culture
Utrecht University

2020-2021

Abstract

The present study investigated the discourse of the Colombian government of Indigenous communities in the context of the data platform of the Colombian government. Considering that data studies are a relatively new academic field, previous research suggested that ongoing power relations are obscured in data collection systems and data sets. However, no earlier research on Colombia's Indigenous people has been conducted. This study used critical discourse analysis to assess two categories of the open data platform in order to better understand the types of datasets that come into each. The analysis suggests that the placement of datasets under specific categories, reveals socio-political discourses of ongoing coloniality that promote inequality and embellish power relations. The findings indicate that the open data platform is now working against the government's goal of increasing trust and transparency among Colombian citizens. Future ethnographic study will create new conversations to better understand this, keeping in mind that each Indigenous group evolves differently.

TABLE OF FIGURES

INTRODUCTION

Over a decade ago, the open data movement started from the idea that data should be free and available for everyone to use without restrictions (Kitchin, 2014). This movement has grown worldwide and governments are starting open data platforms to make data available for citizens, companies, researchers, and anyone who wants to use it (Dawes et al., 2016). Colombia joined the trend and developed a promising open data platform that has gained international recognition. (Constain et al., 2019). Colombia has entered a new phase of technological advancements (Radway, 1980) and wants to reach communities across the country. They intend to use this information to develop and improve public services and people's lives. As useful as it may appear, the usage of open data platforms may hurt citizens, particularly those who do not have access to government services. Sometimes, this data does not represent reality, and often it is categorized in a way that reinforces social discursive practices (Bowker and Star, 1999).

The ambiguity of data systems may create opportunities for Indigenous people to speak and advocate for their communities on the one hand, but it may also generate a cycle of marginalization and misrepresentation (or no representation at all) on the other. Indigenous communities could benefit from data systems to advocate for their rights (Davies et al., 2019). However, it is also the Colombian government's job to enhance this system so that it benefits all Colombian citizens. It is critical to develop a representative ontology and knowledge reproduction. As a result, the focus of this study will be on the Colombian government's discourse on Indigenous people in the context of its data platform. It's critical to examine the government discourses found in the data categories, as they may lead to tangible creations that don't meet people's demands. This research will be a complex understanding of Colombia's decolonization history, but also about how the open data platform of the government reinforces power relations. The data structure reveals that the government has a

colonial discourse. If I can raise awareness about the importance and impacts of datasets, data collection, and categories in open data platforms as a tool to define Indigenous communities' identities through this research, I will be able to enhance their lives by ensuring that future public services are more inclusive. It may be possible to avoid undermining Indigenous identity through data collection and platform data categories.

Although the effort of the government to create transparency through a data platform is valuable for Colombian people, it is creating some other issues regarding media representation. I hope to demonstrate how data categories reinforce Indigenous groups' discursive practices in Colombia through this study. I aim to illustrate that the government misplaces "communities' needs" under a category that favors foreign affairs by analyzing two categories of Colombia's open data platform: "Social inclusion and reconciliation" and "national mapping.". In addition, I aim to demonstrate how the government's interests are obscured within the data system, weakening Colombian citizens' efforts for transparency. Overall, the goal of this study is to show how the categorization of datasets reveals socio-political discourses of coloniality that promote inequality and embellish power relations.

THEORETICAL FRAMEWORK

Data studies have opened discussions around concepts such as "big data" and its impacts on society and culture. The field is still broad, and as we move toward a more data-driven world, new themes and dialogues have emerged inside academia. There have been more discussions about politics, democracy, citizen participation, mobility, and other related themes. The fundamental concepts and views that will be employed in this study will be expanded upon in this theoretical framework. It will analyze the key elements of this study to understand *what is the discourse of Indigenous communities in the context of the open data platform of the Colombian government?*

Who are the Indigenous Communities?

Indigenous communities, also known as first peoples, native peoples, Indigenous peoples, tribal peoples, and so on, do not have a universal definition. The list continues to increase, and there is no agreement on what these people are or represent to society. The United Nations and the International Labour Organization (ILO) recognize that there is not a universal definition. But, under Convention No. 169, some criteria define them as "Indigenous and Tribal People" Jackson (2007).

Figure 1.0 Criteria ILO - Indigenous peoples

|	Subjective criteria	Objective criteria
Indigenous peoples	Self-identification as belonging to an indigenous people	Descent from populations, who inhabited the country or geographical region at the time of conquest, colonisation or establishment of present state boundaries. They retain some or all of their own social, economic, cultural and political institutions, irrespective of their legal status.
Tribal peoples	Self-identification as belonging to a tribal people	Their social, cultural and economic conditions distinguish them from other sections of the national community. Their status is regulated wholly or partially by their own customs or traditions or by special laws or regulations.

Although these criteria are important to consider, there has been international discussion about whether these initiatives take into account the complexities of human history in order to develop appropriate policies. In 2000, Mauro and Hardison reviewed international law and policy concerning the rights of Indigenous peoples. They claimed that while the definition of Indigenous communities includes some components of a widely accepted

international framework, it prevents people from making political decisions based on those elements. For example, according to Mauro and Hardison (2000), the present concept of Indigenous peoples comes from a long history of colonization, genocide, and ethnocide, such as that which occurred in the Americas. On the other hand, Indigenous societies in Europe and Asia have a long history of being marginalized by other Indigenous societies. Therefore, there are only certain elements accepted at an international level that define Indigenous communities. These are self - identification, descendants of a specific territory, shared language and culture, possession of common land, differentiation from national cultures, and others (Mauro and Hardison, 2000). However, this description is based on a western perspective in which Indigenous peoples lack political engagement and must battle for their rights year after year.

In his article, *The Idea of Indigenous People,* Beteille (1998) explains the origins of the term "Indigenous people". Similar to Mauro and Hardison, Beteille explains that the idea of Indigenous comes from the need to have a recognized term for international usage rather than to give participation to such communities. After the 1950s, anthropologists started becoming familiar with communities. With a growing discipline, they needed to have a concrete definition of the term "tribes" (Beteille, 1998). There has always been a persistent search for definitions within intellectual disciplines throughout history, but it has resisted enacting legislation to preserve what is defined under these categories. Beteille (1998) explains that the idea of Indigenous people comes from the relationship between them and the territories they inhabited. In the early '20s Indians and Africans were called "Native" and it referred to a "man of color who carried his identity as a native no matter where he went or what he did" (Beteille, 1998). He argues that anthropological academics created a relationship between race and culture with the term "Indigenous Peoples". Despite the long discussions of the term through history, it is still unclear the direction of the recognition and

definition of such groups. Mauro and Hardison (2000) argue that the UN defines Indigenous people to interpret the obligations of external colonies since in 1948 the UN Charter recognized the 'free pursuit and 'self-determination' of non-self-governing territories. It is a topic of debate among academics and international legislators. As a result, the concept of Indigenous people remains ambiguous, and for the purposes of this study, we shall use the "definition" that best fits our corpus. This research will use the term "Indigenous Communities" translated from the Spanish name "Comunidades Indigenas" recognized by the Colombian Government. Different from the ILO, according to the Colombian legislation, "The Indigenous Communities are the human group that lives by the forms of relationship with the natural environment in which the different aboriginal groups settled since before the conquest and have conserved and energized it throughout history". We can already see elements of the government's engagement with Indigenous communities based on this description. Apart from that, when analyzing datasets, this study will only analyze communities that match the country's recognized list. The ONIC, CECOIN, and GHK, in 1995, conceded that in Colombia there are 81 Indigenous communities composed of different ethnicities. This list can be found in the Minister of Culture documents. Figure 1.1 shows a sample of what this document contains including a map with the communities.

Figure 1.1 Indigenous - Communities of Colombia
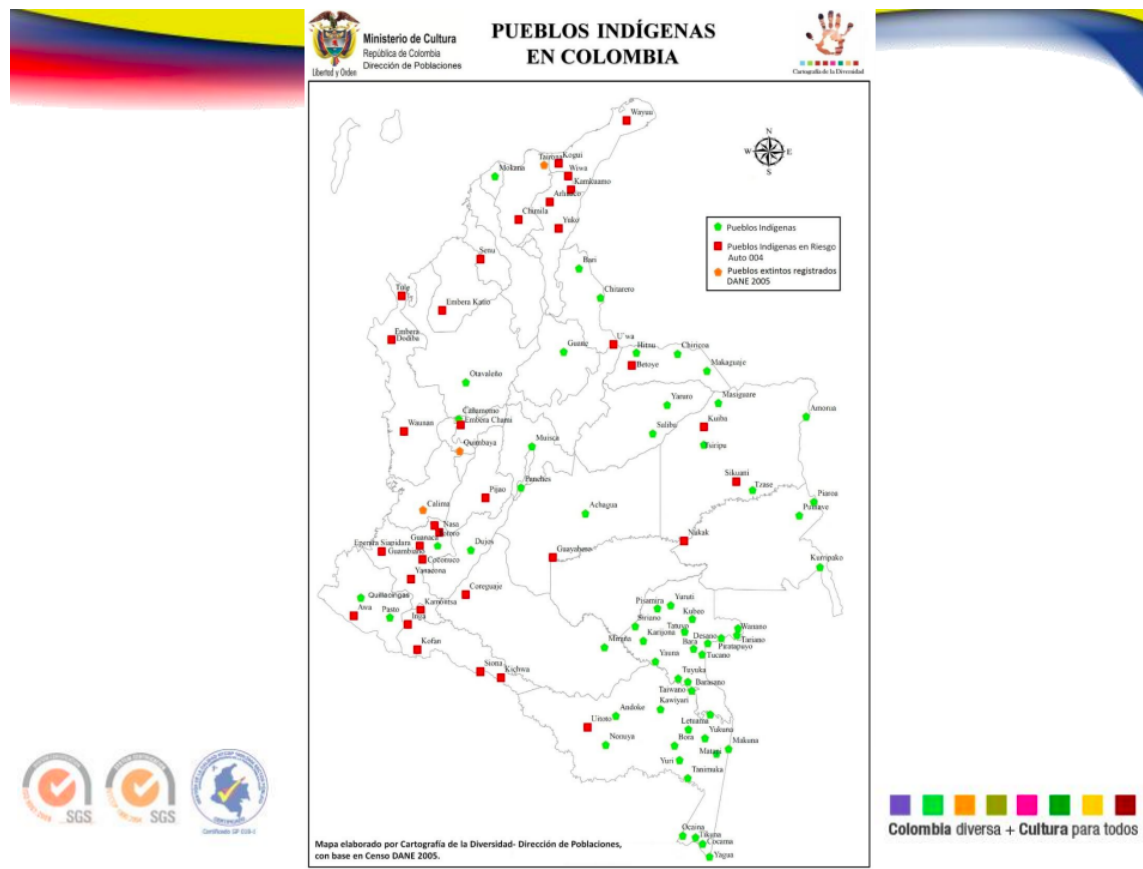
So far, this theoretical framework has clarified the concept and criteria of "Indigenous Communities". A greater knowledge of data infrastructure and how it influences open data platforms is required to understand how data categories strengthen government inequality.

Are Data Infrastructures like bridges?

In 1933, the construction of the iconic Golden Gate Bridge began. As of today, 523 million dollars have been invested and the structural planning and engineering required have amazed thousands of tourists that come every year to see such a megastructure (Schrank, 2004). The project required more than just materials and labor; it also required an infrastructure that would allow one of the world's longest suspension bridges to be built. This included physical systems and services provided by the city to enable productivity in the economy. But how does this relate to data? We often think of infrastructure as a tangible form of service. But just like the Golden Gate Bridge, data also holds a less tangible infrastructure.

Dodds and Wells (2019) conceptualized data infrastructure in a form of analogy with physical systems of construction or also known as "traditional" infrastructure. For instance, with the bridge example, such infrastructure required planning and designing. Data infrastructures also need to be designed to support a variety of uses and "it is a public good that enables the creation of a wide range of products and services" (Dodds and Wells, 2019). Just as the Golden Gate Bridge became part of the Pacific Highway road to help people navigate to a destination, data helps us to navigate to a decision. Dodds and Wells (2019) suggest that by looking at data platforms in this way, we can better understand how data requires more than just connections and physical assets to be maintained, since it allows us to think of data as more than just how it is stored and collected. Therefore, they proposed the following definition:

Data Infrastructure consists of :
1. *Data assets, such as datasets, identifiers, and registers.*
2. *Standards and technologies used to curate and provide access to data assets.*
3. *Guidance and policies that inform the use and management of data assets and the data infrastructure itself.*
4. *Organizations that govern the data infrastructure.*
5. *The communities involved in contributing to or maintaining it, and those who are impacted by decisions that are made using it.*

As Dodds and Wells (2019) point out, understanding data's technological assets and rules, as well as the organizations and communities participating in its contribution and upkeep, is critical. This perspective is different from other academics who have considered data structures more as networks of information.

Gray et al. (2018) examined data infrastructure literacy to analyze the different uses of this concept. They examine that in information policy, data infrastructure is often used to refer to the development of large-scale technical systems of the creation, processing, and distribution of information. This perspective of the data infrastructure is mostly used when

analyzing networks rather than social constructs. For instance, USA President Clinton used a "National Information Infrastructure" to understand the nation's datasets distribution which was also called *networking the nation* (Gray et al., 2018). Among researchers, it is common to understand data infrastructures as a series of connections or relations. Star and Ruhleder (1996) proposed to understand information infrastructures in terms of "relations" rather than as "things". According to this viewpoint, data platforms are continually altering relationships among databases, software, committees, interfaces, and other components. It's critical to understand these linkages in order to realize how they interact. Unfortunately, this viewpoint ignores the fact that data does not only designate aspects of reality. As a result, cultural values have been filtered through data, and this suggestion favors one method of data interpretation over another (Gray et al., 2018). Data can also be multivalent and used by a variety of actors. The authors before them attempted to comprehend the concept of facts in order to show how things are. There is a belief that data and its infrastructure are static, but there are greater opportunities in adopting a concept of data that allows us to investigate social, historical, political, and cultural processes (Gray et al., 2018).

This notion is introduced by Bowker and Star (1999) and it is called "Infrastructural Inversion". This concept allows for the research of data as a quantification element, but also of data as a social practice. Understanding data structures at both a technical and a social level is critical for this research because it allows us to not only understand the patterns resulting from discursive practices in the data collection process, but also to align them with the government's socio-political structural organization. Bowker and Star (1999) would allow us to create a link between the development of statistics and quantification with the creation of governing populations (Gray et al., 2018). Bowker and Star (1999) support this study because it allows for the establishment of a relationship between the quantifying component of accessible data and the governing that leads to discursive practices. Bowker and Star

(1999) allow this research to build "infrastructural imagination," which means they perceive data as dynamic links that display various ways of viewing data from a social perspective that would be more appealing to the general audience. As discussed earlier, data infrastructures are not as static and rigid as a bridge might seem. Therefore, considering the dynamic nature of data that is entwined within the social and technical aspects will allow us to understand how discursive practices from the government are present on the open data platform. In addition to Bowker and Star (1999), Ruppert (2015) has a similar point of view. He describes that data publics are constituted by "dynamic, heterogeneous arrangements of actors mobilized around data infrastructures" (Ruppert, 2015). He sees public data platforms as diversity brought about by a mix of information regulations and civic technology (Gray et al., 2018). This viewpoint provides practical methods for comprehending civic engagement. However, in order to conduct this study, we must establish links between discourses and open data structures.

Different problems about the role of governments and public institutions in the interchange and creation of data platforms have arisen as a result of data infrastructure. Bowker and Star (1999) and Ruppert (2015) allow us to see data not just as a dynamic but also as a social construct, and their perspectives idealize, in some ways, how data platforms may or should work in a more "developed society". For example, Bowker and Star (1999) use data examples mostly from medical systems and medical bureaucracies such as the international classification of diseases, nursing interventions, or even the classification of viruses. Although there is a suggestion to comprehend certain aspects of data at a political level, the idea that structures are solely the result of political internal actors does not allow for data understanding in the context of a growing country like Colombia. Another example is that in most of the nations cited by these writers, a body or organization is responsible for collecting and cleaning data. In a country like Colombia, however, a lack of budget,

organization, and efficiency results in an inconsistent data system that is governed by no one. Indigenous populations in Colombia frequently lack access to public health services, making them undetectable in this data. Indigenous communities, on many occasions, are excluded from the government's data platform and basic technological systems such as the identification system, or the census data. Therefore, this study agrees with Bowker and Star (1999) on data infrastructures, but it needs to be supplemented with a perspective that includes more elements from a developing country. Dodds and Wells (2019) raise this problem by claiming that most road building in developing nations is funded by foreign investment, raising the question of who benefits most from this type of infrastructure. Therefore, Dodds and Wells (2019) say we need to understand who controls our data infrastructure, but also "recognize that there is an evolving variety of governance models. The right model for the governance of a given element of data infrastructure might vary across sectors, nations, or communities" (Dodds and Wells, 2019).

Critical Discourse Analysis in Media Studies

Finally, we must comprehend Critical Discourse Analysis (CDA) in order to comprehend government inequity concerning indigenous populations that are present in the data platform of the Colombian government. This term has been used in a variety of fields, but this theoretical framework will emphasize the use of the term within the context of media studies.

According to Ramanathan and Tan (2015), Teun Van Dijk, Ruth Wodak, and Norman Fairclough have been the main contributors to this field. In general, they all understand CDA as a linguistic orientation that seeks to analyze social and political actions, but they have different methodological approaches. Van Dijk, for instance, believed that "CDA is a proposition which focuses on how power abuse, dominance, and inequality are practiced in the discursivity of the social and political context" (Ramanathan & Tan, 2015). As a result, he

sees CDA as a method of analyzing social phenomena that necessitates a "multi-dimensional approach." Wodak, on the other hand, sees CDA as a method for analyzing social phenomena that necessitates a "multi-methodological approach." Van Dijk states that the mental representation of a group depends to a large extent on social structures. Van Dijk claims that because CDA focuses on text, this type of social behavior influences how people offer or focus information in a text. As a result, he feels that CDA examines social issues in great detail. However, Wodak and Van Dijk were not the only ones who saw CDA in this way. For example, Janks (1997) defined CDA "as a form of social practice whereby critical theories are applied to analyze opaque relationships" (Ramanathan & Tan, 2015).Janks, like Van Dijk, has studied CDA as a means of comprehending language with strength. But how does CDA fit into media studies?

Phelan (2017) highlights the importance of this concept in the field of media studies. He mentions that it was not until Fairlough's approach that debates and different perspectives arose in media and communication studies. Fairclough (1995) argued that CDA aims to "systematically explore the opaque relationship of causality and determine between a) discursive practices, events, and texts, and b) wider social and cultural structures, relations and processes" (Ramanathan & Tan, 2015). He "framed his approach to media discourse as a development of the semiotic methods of cultural studies scholars" (Phelan, 2017). With this approach, CDA has been used to analyze how media representations modulate discursive constitutions of different social phenomena. Phelan (2017) identified two different types of researchers within the space of CDA and media. Some are linguists that analyze media and others are media scholars that apply CDA theories and methods. This study is an example of the last. It will benefit from CDA's ideas and methods for understanding social phenomena such as indigenous societies' discursive practices. Jorgensen & Philip (2002) mention different features that explain why this method is the most developed for communication,

culture, and society. First, they explain that the character of social and cultural processes and structures is partly linguistic discourse. In other words, although CDA reveals more information on how linguistic-discursive dimensions change social and cultural phenomena, they argue that discursive practices in everyday life also reveal social and cultural changes that are not linguistic. CDA, according to Jorgensen and Phillips (2002), is a "multi-disciplinary approach to understanding the link between discourse and socio-cultural processes." This allows revealing perspectives that may not be inherent in the text in the field of media studies. Second, Jorgensen & Philip (2002) explain that discourse is both constitutive and constituted. They mention that when analyzing discursive practices in media, researchers need to take into consideration that these practices shape new forms of politics and influence society, but also that discursive practices are influenced by other social forces that do not have a discursive character. For instance, in countries like Colombia, there is a lack of resources and funding for data technology which makes it difficult for the government to focus on specific issues of the open data platform. This could lead to the government unwittingly reproducing certain discursive practices. In Colombia, however, the government has attempted to digitize key areas of our daily life, such as service payments, financial aid requests, and doctor's appointments, among other things. These new digital experiences were not feasible ten years ago, but they now lead to political discourses about the country's "progress" and "development". As Jorgensen and Philip (2002) pointed out, it is critical to investigate not just how other social forces may impact discursive practices, but also how media generates new political discourses in this study. Finally, Jorgensen and Philip (2002) emphasize that because CDA's naturalness is to show unequal relations in the role of discursive practice, there are ideological effects in it. However, they might contribute to creating and reproducing unequal power relations between social groups. Because the media has a two-way interaction with its citizens, this is an important part of our study. On the one

hand, media seeks to meet people's requirements; on the other hand, media tries to generate and spread certain messages to citizens. What this research will focus on is that media is not an actor by itself but that it is shaped and built with a set of actors, entities, and in general with an infrastructure.

METHOD

Given that Indigenous Communities do not have a universal definition, this study will investigate social and political acts using Critical Discourse Analysis (CDA). As previously said, CDA's approach will be centered on media studies; as a result, the relationship between discursive practices, socio-cultural structures, and data platforms will be evaluated in this study. As a media scholar, I propose to use CDA's theories and methods in the following context to study the discursive practices of the Colombian government in the context of the data platform of the Colombian government. So, it is important to go over the corpus that was chosen for this study first.

Open Data Categories

Out of the 26 categories of the open data platform, this research will focus on two: the first is referred to as "Social Inclusion and Reconciliation" in Spanish or "Social Inclusion and Reconciliation" in English. The second category is known as "Mapas Nacionales" or "National Mapping" in English. From now on we will refer to them by their name in English. These categories were selected under the next criteria:

1. There are 15 or more datasets under the category.

2. 3 or more of the datasets of the category correspond to Indigenous matters or have the word "indigenous", or correspond to an Indigenous community.

3. They provide significant examples of possible discourses that are present within their datasets.

There are 78 datasets under "National Mapping," and 190 datasets under "Social Inclusion and Reconciliation". The datasets will be filtered to discover which ones correspond to Indigenous communities, and the sample that "represents" the communities for census reasons, territorial residence, or political engagement will be chosen. What this research will evaluate is the type of datasets that fall into types of categories (only 2 of them).
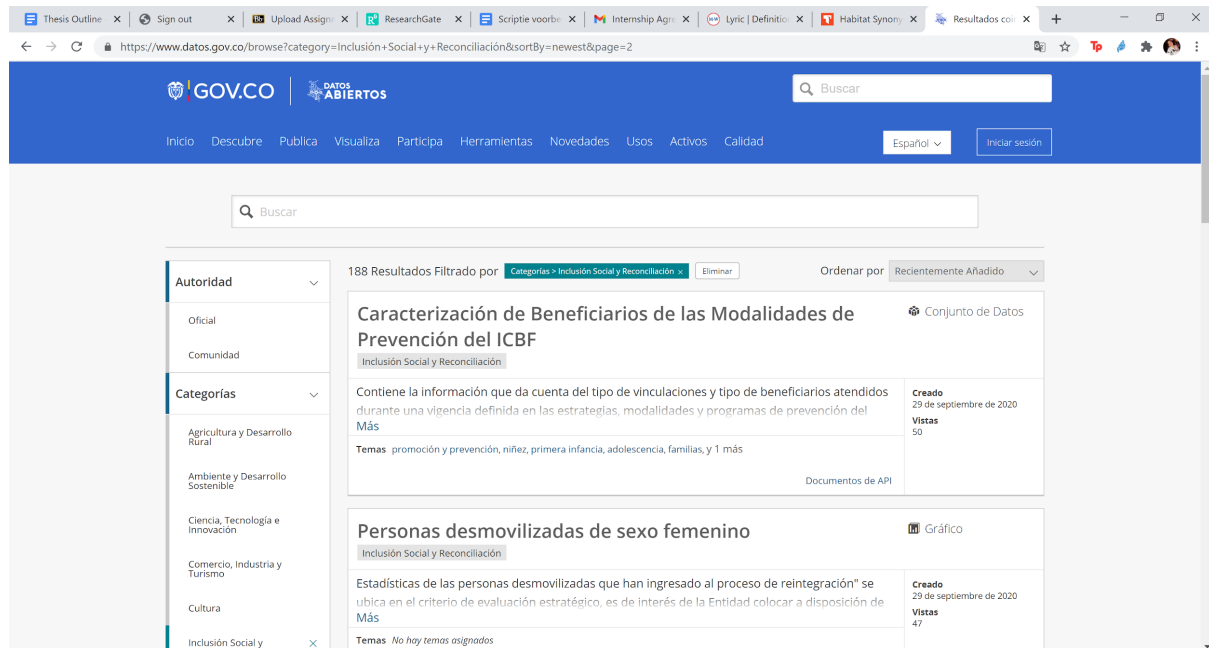

Figure 1.2 The open data platform: https://www.datos.gov.co/

Critical Discourse Analysis In The Context of the Colombian government's data platform

This research will focus on the perspective of CDA from Norman Fairclough. One of his most important characteristics is to focus on the "Investigation of change". It relates to the study of intertextuality (textual relationships) since it allows us to comprehend changes in speech across time and from a linguistic standpoint.

" Discourse is an important form of social practice which both reproduces and changes knowledge, identities and social relations including power relations, and at the same time is also shaped by other social practices and structures" (Jorgensen and Philip, 2002).

I propose CDA for this research because it will allow me to create the link on how the Colombian government's data platform assimilates with the social conditions of the country. More crucially, Fairclough's approach to media discourse will allow me to examine how Colombia's categorization system works as an epistemic tool, leading to the government's discursive practices. Fairclough's three-dimensional model has three components to understand critical discourse analysis: text, discursive practice, and social practice. The text portion of CDA will mostly consist of visual images and datasets acquired through a computational system (Tableau, Canva) to represent the textual categories of a collection of datasets belonging to two separate open data categories. The contrast of visualizations and data will enable me to examine discursive practices among one minority group in Colombia: Indigenous Communities.

Finally, I will look at the historical and social context of the Colombian government's data platform to understand the discursive practices. I'll look at policy documents to learn more about the platform's motivations and development. The open data platform of Colombia is used to improve and create public services that claim to benefit Colombian citizens. Government discourses in the context of the data platform of Colombia are already problematic for "regular" Colombian citizens. But, with this research, I want to evaluate the discourses that affect Indigenous communities. It's vital to note that, while Indigenous peoples are Colombian citizens, their current legal and social standing is impacted differently than that of "ordinary" citizens. This will aid us in comprehending how accessible data is affecting public discourse in such a way that it appears natural how Indigenous Communities "should" live. This allows us to examine the production and consumption of textual elements, which includes researchers, private and public bodies, and everyday information seekers. I will analyze what the creator of the database (Colombian Government) suggests about it and what this implies to Indigenous communities in the country.
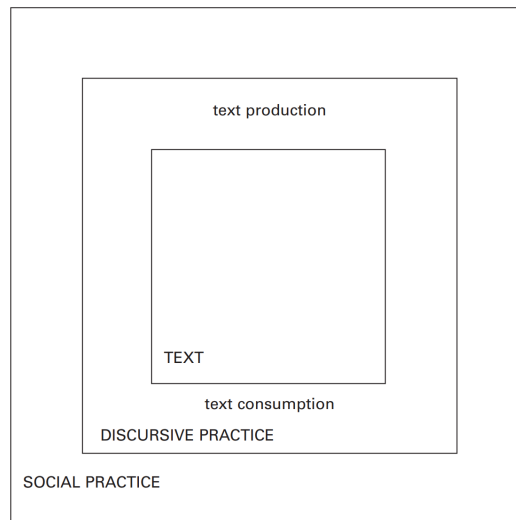
68 DISCOURSE ANALYSIS AS THEORY AND METHOD

```
┌─────────────────────────────────────────┐
│                                         │
│   ┌─────────────────────────────────┐   │
│   │        text production          │   │
│   │   ┌─────────────────────────┐   │   │
│   │   │                         │   │   │
│   │   │                         │   │   │
│   │   │                         │   │   │
│   │   │                         │   │   │
│   │   │                         │   │   │
│   │   │ TEXT                    │   │   │
│   │   └─────────────────────────┘   │   │
│   │        text consumption         │   │
│   │ DISCURSIVE PRACTICE             │   │
│   └─────────────────────────────────┘   │
│ SOCIAL PRACTICE                         │
└─────────────────────────────────────────┘
```

Figure 3.1   *Fairclough's three-dimensional model for critical discourse analysis (1992b: 73)*

Figure 1.3 Fairclough's three-dimensional model for critical discourse analysis

DIGITAL GOVERNMENT

In 2018, the Colombian government released act 1008 to implement *the politics of the digital government* (Constain et al., 2018). This initiative was led by the Ministry of Information Technologies and Communications (MinTIC) and founded by the Organization for Economic Co-operation and Development (OECD). The objective is to "promote the use and exploitation of the technologies of information and communication to consolidate a state and Colombian citizens that are competitive, proactive, and innovative; this to generate public value in an environment of digital trust" (Constain et al., 2018). If Colombia continues to make more government data openly available and encourages its reuse, it will position itself as one of the leading countries in the field of open data (OECD, 2017). In Colombia, the law 1712 of 2014 regarding transparency and access to public national information defines open data as "all primary data or raw data, which are in standard and interoperable formats,

that facilitate its access and reuse; these are under the custody of public or private entities that fulfill public functions and are made available to any citizen freely and without restrictions so that third parties can reuse them and create services" (OECD, 2018). The main feature of Colombian open data is that it is published on websites with free access and is available for use by any citizen (Constain et al., 2018). The MinTIC adopted the principles and good practices of data recognized by the International Open Data Charter (Constain et al., 2019). This includes, as illustrated in figure 1.4, the following:

- Data must be open since it is first generated.

- Data is relevant only if it contributes value to the user of the information.

- It is free and accessible.

- Can be compared from different departments, territories, and time frames.

- It strengthens governance and transparency.

- Allows building new knowledge to improve social and economical value.

Figure 1.4. Principles of open data

To ensure that citizens could benefit from open data in an environment that adhered to these principles, the government established the TIC for *Citizens* and the TIC for the *State*. (figure 1.5) (Constain et al., 2018). So, according to Constain et al. (2018), the country could benefit from good data use and practice in the following ways:

1. Transparency and Social control: publishing information regarding the tasks and jobs of public entities creates investigation processes made by academics or journalists. This allows us to monitor how the public money is used and invested.

2. Improvement and creation of products, services, and innovative companies.

3. Predict and prevent phenomena such as health, crime, and environment.

4. Generate new knowledge.

Figure 1.5 Components of open data

"The Colombian government will focus on digital government as the driver of competition and strengthening the Colombian market, as well as building up the links of trust between the

public and State" (OECD, 2018). For this, since 2014, Colombia's constitution recognizes access to information as a fundamental right for its citizens and created five key elements that compose successful open data governance (Constain et al., 2018). This includes norms and an ecosystem of actors, tools, incentives to monitor data. The ecosystem of actors, as depicted in Figure 1.6, is made up of public entities, journalists, academics, citizens, and developers (Constain et al., 2019). Colombia aims to create an inclusive platform to foster trust in this new technological era (OECD, 2018).

Figure 1.6 Ecosystem of actors of open data.

Colombia's legislation regarding data already existed, but it wasn't until the implementation of the *digital government* that Colombia started to look forward to a more digitized country (Constain et al., 2019). Colombians are becoming more aware of the significance of open data as a concept, as well as the benefits and improvements it could

bring to the country. Most importantly, the government has a tremendous intention for journalists and academics to investigate such data to rebuild trust with its citizens (Constain et al., 2018). "The increasing use of data sources [...] in local government policymaking raises questions about the quality of these datasets and how they represent the communities impacted by their use" (Dodds and Wells, 2019). The government is attempting to regulate and create a good practice of these data by providing a new set of laws, rights, and making information available. While the new *digital government* initiatives appear to be promising for the country, other issues are emerging that could undermine Colombia's tremendous efforts in the post-conflict era.

HISTORICAL BACKGROUND

In 2016, Colombia's government signed a historic peace agreement which was an effort to end a period of war with the illegal armed group FARC (Ley de Justicia y Paz, Ley 975). During the four years of negotiations, Colombia had tremendous international support. Some of the entities that finance the postconflict fund are the World Bank, the Swedish International Development Cooperation Agency, the International Cooperation for Migration, and the United Nations High Commissioner for Refugees (UNHCR)(Trindade, 2009). A new post-conflict era began, and the government began to implement the new socioeconomic changes that had been agreed upon. The government established the Special Jurisdiction for Peace (JEP) to ensure that all parties involved speak the truth about the years of conflict without fear of being prosecuted under Colombian law (Beittel, 2015). In addition, the government started a number of initiatives to heal the country and reintegrate thousands of FARC members into society (Guasca et al., 2021). Some of these initiatives are reflected in the open data platform, which has created a new set of categories and gathered data that is relevant to the post-conflict era. In this scenario, we find most of the datasets that have information regarding subsidies, government aids, and social plans created for the ex-FARC

group members. Nevertheless, 10,5% of the victims were either Afro-Colombian or Indigenous Communities who belong to the lowest economical *status* (Rettberg, 2019).

INCLUSION OR INTERNATIONAL RECOGNITION?

With 75,000 people attempting to reintegrate into society, it is evident that the Colombian government would require information from ex-guerrilla members and thus created a new category on their platform (Beittel, 2015). As explained by Bowker and Star (1999), classifications are overlapped in our lives and sometimes governments use bureaucratic classifications to run organizations that will become more visible and contain important information about us. In this case, an example of a classification that was created to run an organization or institution in Colombia is the category "Social inclusion and reconciliation". It is one of the categories where we can find the majority of the datasets and information about the peace process. What this category did not do was collect data from all of the other 6,453,000 victims of the country's 60-year conflict (Rettberg, 2013).

There were three Indigenous Communities datasets in this category, with only one of them being accessible. I could only access a dataset called in Spanish "priorización indígena para el programa de subsidio económico otorgado por Colombia mayor en el municipio de Sincelejo"; translated to English it is "Indigenous priority for the economic subsidy given by Colombia in the Municipality of Sincelejo". From now on, I will refer to it as Dataset A. When analyzing this first dataset, we can see that most of the data collected are regarding Indigenous members who are awaiting a government subsidy in this region of the country. Most of the subcategories of Dataset A contained information about their identification: ID number, address, birth date, and so on. But the Dataset also gathers information about their health score which is based on Colombia's resolution number 630 of April 21 of 2020. It assigns residents a score depending on a number of factors, including whether they are covered by the minimum health plan, whether they are part of a UNESCO-declared

manifestation of cultural patrimony, and whether they live in a rural portion of the country, among others. In the following section, I will explain why Indigenous Communities' economic and health data is categorized as "Social Inclusion and Reconciliation."

Figure 1.7. Social inclusion and reconciliation dataset

Data Production and Consumption

The open data platform of Colombia is a collection of databases and information gathered by different entities of the government. This information is cleaned by the Ministry of Information Technologies and Communications (Min TIC) before it is published on the open data website (Constain et al., 2019). A description of who provided the data and who is responsible for its publication is included with each dataset. Although the Presidency of the

Republic administers the majority of the data collected for the category "Social Inclusion and Reconciliation," the information containing datasets of Indigenous communities is gathered by the Municipality of Sincelejo and provided by the secretary of social development (Figure 1.8). Dodds and Wells remind us that "the right model for the governance element of data infrastructure might vary across sectors, nations, or communities" (Dodds and Wells, 2019). Although the MinTIC has specified specific requirements for data collecting, the vast majority of the open data platform's datasets lack collection patterns. As a result, the disparity in their data collection system corresponds to the substantial disparities in the platform's datasets. The main concern is that this 'disparity' system of data gathering has become normalized among the institutions involved. Once a system is in place, the practical politics of these decisions are often forgotten and often turn into software such as the open data of Colombia (Bowker and Star, 1999).

Figure 1.8 Data Flow -Dataset A
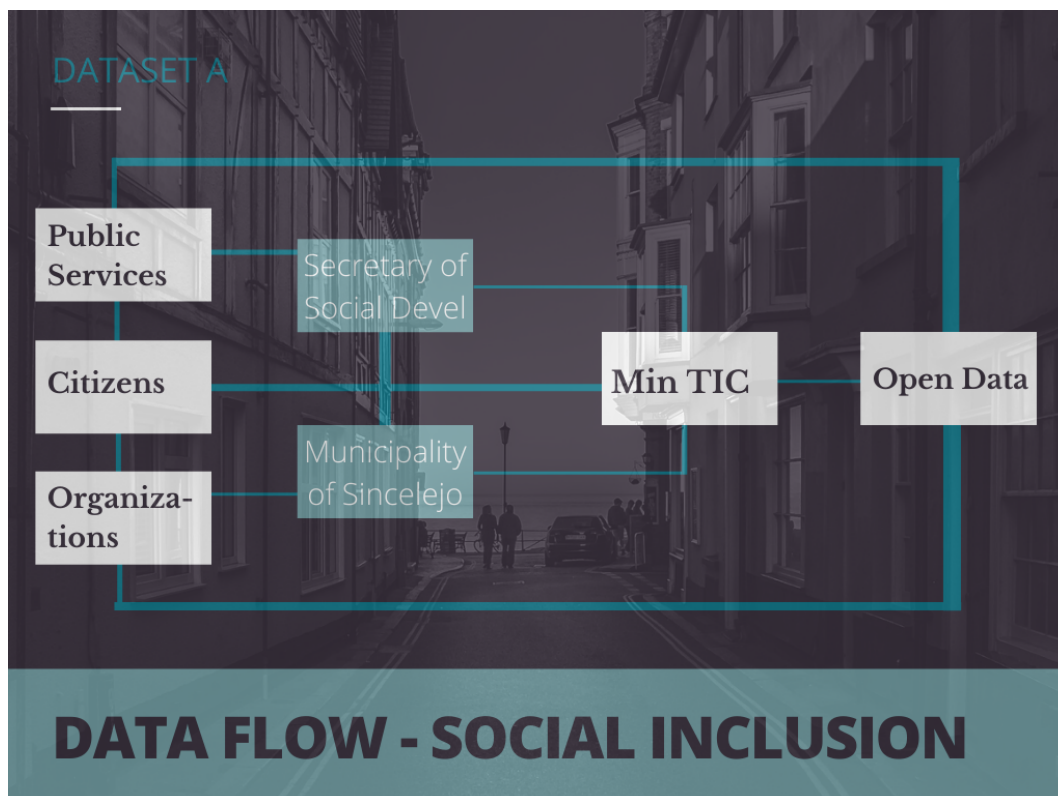
We wondered at the start of this section why information about Indigenous Communities' economic and health status would be placed under the category "Social Inclusion and Reconciliation". With this question, I intend to address Bowker and Star's idea that humans stand for the most part in formal ignorance of the social and moral order of categories created by these invisible, potent entities. I will disclose some of these social and moral orders that Colombian citizens wouldn't be aware of otherwise. Through this research, we cannot say for certain that there is a clear intention of misplacing "community needs" under specific categories, but the lack of regulation regarding the data system of Colombia aligns with the lack of caring and attention for Indigenous Communities.

There was an opportunity for a more inclusive society after the peace agreement was signed. However, "conflict issues are linked to deeply entrenched formal and informal institutions which partially explain ongoing polarization and widespread contestation" (Rettberg, 2019). Institutionally, there is a set of categories created for mostly ex-guerrilla members but it excludes information of the 20,000 indigenous victims of the conflict ("Colombia Reconoce", 2020). I want to highlight not only the importance of data platforms but also to identify why it is important to emphasize governance and legislation to create equitable access to the benefits of data (Dodds and Wells, 2019).Similar to Bowker and Star (1999), I demonstrate how these categories are created, as well as how they remain invisible to humans, such as indigenous communities, resulting in exclusion and violations of their rights. Although it is debatable whether Indigenous peoples should be included in the creation, classification, and categorization of open data, classification does play a role in perpetuating institutional and citizen discourses.

First, compared to all the other 187 datasets, the data regarding Indigenous Communities is gathered by a local entity rather than by the Presidency of the Republic. In

the 1950s, the UN recognized that universal human rights were not enough to protect indigenous people from persecution, genocide, and discrimination (Mauro & Hardison, 2000). This institutional structure of data collection shows how Indigenous Communities face significantly less government presence than other citizens and historically have less engagement with the government (Anderson-Smith, 2008). The government constructs a narrative that Indigenous issues can only be managed by local institutions and do not require intervention from higher levels of government. The problem is that when these issues come mainly from political institutions, they become invisible through data systems. "Politically and socially charged agendas are often first presented as purely technical and they are difficult even to see" **(**Bowker and Star, 1999). Second, the dataset presented under this category has nothing if very little to do with social inclusion and reconciliation. The data was collected by the secretary of social development and collects information regarding economic subsidies and health status. There are several other categories in which this data could have been placed: social development, health, and social protection, economy, and finances, among others. This shows that layers of classification might become unfolded into a working infrastructure. Now, we see the original political intervention and prevent naturalization of political categories **(**Bowker and Star, 1999). Collecting data on people's health, economic situation, and prioritizing scores clearly has little to do with "social inclusion and reconciliation," which brings us to the final argument. Because the government needs to provide measurable outcomes to international entities, information about Indigenous Communities is included in data categories such as "Social Inclusion and Reconciliation." "The need for significant investment can also unintentionally exclude some organizations and communities from participation in the creation of data infrastructure" (Dodds and Wells, 2019). Therefore, the needs of the Indigenous Communities are only relevant for the government when international organizations are willing to invest in the country, but those

are not addressed adequately. As previously analyzed, the definition of indigenous communities is somewhat collective because lawmaking tries to promote the authority and decision-making of Indigenous in their territories (Mauro & Hardison, 2000). "One of the major objections to the novel rights of indigenous peoples has been that they are largely rights of collectivities, not individual rights" (Wiessner, 2011). Such classification creates a legal regime of autonomy but also places indigenous in a framework of international laws to define their living conditions and vulnerabilities. Categories of the open data classify indigenous creating a quota system that sets the limitations of the indigenous autonomy (Bowker & Star, 1999). Because the information is not categorized under the matters to which it should belong, this categorization does not represent the community's needs. This means that services created for this matter might not represent or use the most accurate data. "The need for significant international investment may lead to the creation of an infrastructure that is less optimal because of the lack of consideration of different or diverse perspectives and needs" (Wells and Dodds, 2019). It will not approach the information with the relevance that it requires and it transforms into a service based on inaccurate discursive practices. One example of a recent historical event in which government discourses are displayed in the Colombian government's data platform is "Social Inclusion and Reconciliation." However, we will go on to another category to better grasp the value of analyzing such dialogue in this new technological era.

WHAT ARE WE MAPPING?

The peace agreement was not only a tool to demobilize thousands of guerrilla members and end the cultivation of illicit drugs, but it also became a transition towards a more structured justice regarding Human rights violations (Rettberg, 2019). Throughout the years of war, Indigenous Peoples have been subjected to a slew of transgressions, resulting in a long history of hostility between them and illegally armed organizations. The agreement

tries "to compensate the more than eight million victims in the country who have formally registered within the Colombian state" (Rettberg, 2019). Crimes and Human Rights violations related to land play a central role in the aspiration of the Colombian transitional justice structure because most of the victims were forcibly displaced (Rettberg, 2019). However, these communities have been subjected to crimes that predate the war. There has been an ongoing battle between the government and the Communities over national territories since the colonization era. "Indigenous peoples raise claims for self-determination and land that vindicate local customs, laws, and ancestral territories" (Rodriguez-Garavito & Carlos-Arenas, 2005).The ongoing fight for natural resources, territorial sovereignty, and cultural variety in the country is a perennial topic of discussion. With this part of the research, I want to address Bowker & Star's (1999) view on the impact of classification such as those of regions, and natural resources that have on public policy and economic decisions. Ecological classifications have a tendency to develop bureaucratically complex processes, as I shall demonstrate in the next section of the study.

National Territory

The second category for this study is "National Mapping", which is defined by data that allows us to draw borders inside a country. For example, the distribution of natural reserves, the placement of specific minerals in various regions, the coverage of utilities such as power and water across the territory, and even the distribution of firemen across the country are all examples of information found in this category. Out of the 78 datasets found in this category, three of them belong to Indigenous Communities information and two had duplicates. All of the data collected is regarding Indigenous living areas, their addresses, and their *etnias* (subcategories of Indigenous Communities). Only one dataset had extra information and it will be used to exemplify discourses found in the context of the Colombian government's data platform. This dataset is called in Spanish "Comunidad Indígena de

Caldas", translated to English "Indigenous Community of Caldas". From now on, I will refer to this dataset as Dataset B.

In the "National Mapping" category, unlike the first, the majority of the data is acquired by local entities. In this case, Dataset B is collected by the governance of Caldas and the data provided was under the secretary of planning. The information gathered will be presented in three parts (Figure 1.9).

Figure 1.9 Information Collected in Dataset B

DATASET B - INDIGENOUS COMMUNITY OF CALDAS

LOCATION	HISTORY	CULTURE
Country sub-region	Population	Ancestral Traditions
Municipality Code	History of Constitution	Dances
Municipality	History of Recognition	Chants
Sector (Rural or urban)		Gastronomy
Vereda - (rural sidewalk)		Handcrafts
Type of via (pavemented, no road)		Auctochthonus games
Geo-reference		

Location

The majority of the information in this collection pertains to the inhabitants of this community. Analyzing the ambiguity of data, it is obvious that it has the potential to provide very particular information about these communities' living conditions, but it also has the potential to contribute to the creation and reproduction of unequal power relations between social groups (Fairclough, 1995). This data could help the government improve Indigenous people's quality of life, which could be one of the reasons why the secretary of planning gathered it. So, why are these types of datasets classified as "National Mapping"? What exactly are they mapping out? With these questions, I want to learn more about Van Schie's perspective on how classification systems interact with technological infrastructures that prioritize information based on different values of different groups, and how those groups become racialized as a result of the violence they have experienced (Van Schie et al., 2020).

Knowing the location and demographics of this community could be an indication of mapping from one perspective. On the other hand, the Colombian government profits from political and cultural oppression, natural resource exploitation, and the determination of Indigenous groups' territorial rights (Hristov, 2010). "Whether or not a region is classified as ecologically important [...] it bears significantly on future economic decisions" (Bowker & Star, 1999). That is why it is brought to the notice of the public that this dataset contains information regarding Indigenous communities' history, as well as data on whether or not this history has been recognized legally or in the constitution. Colombia is still under international pressure, as previously stated, and the government will be unable to act in areas designated by UNESCO as Human Patrimony. The goal of charting Indigenous territories may be to enhance their lives, but the community's relationship with territory and history goes beyond that.

History

The Inter-American Commission of Human Rights recognizes that the Colombian Government is instituting a system of national Indigenous parks and reservations to recognize the claim that Indigenous communities have over their traditionally occupied territories (Wiessner, 2011). The government, on the other hand, is attempting to simply engage Indigenous Peoples in the discussion before exploiting their lands. "The 1991 constitution and subsequent legislation specify that "customary law" will have the power within Indigenous territories" (Jackson, 2007). This means that social and economic norms that are acknowledged as mandatory rules of behaviour, practices, and beliefs will be treated as if they were laws. To put it another way, this law will greatly benefit Indigenous practices, but the government will have the last say on whether or not these customs are part of the social and economic system. The real issue is that through technological conditions of production, the government is building a classification system and producing the main standard and abstract argument about the representation of Indigenous Communities (Bowker & Star, 1999).

There are presently 302 reservations, involving a total of 26 million hectares of land that is inhabited by 310,000 Indigenous persons. Since 1987, Colombian law guaranteed Indigenous communities their right to exploit the renewable natural resources on those lands (Roldan Ortega, 2004). But, "as a result of the international financial crisis, the international price of gold in Colombia doubled in just a few years" (Idrobo et al., 2014). The exploitation increased in these regions and the business became very profitable. "The increase in the illegal exploitation of metals like gold has exacerbated violence in municipalities with an abundance of such minerals" (Idrobo et al., 2014). Classification systems often lead to political and social struggles, but these are difficult to approach when there are clear unequal power relations (Bowker & Star, 1999). The government failed to protect indigenous

territories but yet classifies them in terms of 'land' rather than as 'individuals' with rights. Indigenous people's mistrust of the government may be reflected in data collection methods. People will no longer consent to data gathering and feel discouraged toward data efforts if there is no trust in how personal data is used in the future (Dodds and Wells, 2019).

Dataset B categorizes whether Indigenous areas are acknowledged as part of 'national' or 'world history,' and whether they are protected and cannot be exploited, based on the information collected. In Dataset A, it was discovered that when it comes to considering Indigenous requirements, Colombia's international recognition comes first because it provides the country with more economic benefits. Dataset B, like Dataset A, demonstrates that unfettered access to these areas is contingent on international recognition. Dataset B, on the other hand, differs in that the reason for mapping these communities is for national objectives such as natural resource extraction. Similar to Van Schie (2018) research, this data could be considered more 'administrative' than about the people themselves. This causes that the government, through ecological classification, normalizes racist ideologies that later become policies. The government uses media not as an actor by itself but shaped with a set of entities and with an infrastructure proposed by them (Fairclough, 1995). They have imposed a classification system in which Colombian citizens and Indigenous communities have naturalized powerful sets of classifications of knowledge (Bowker and Star, 1999). Indigenous tribes have normalized a set of constructs and turned them into "administrative" data, despite the illegal exploitation of their territory. When people lose their value as individuals, a rigid system of racial segregation and unfairness emerges, which will continue to reproduce unless categorizations become more egalitarian and inclusive.

Colombian legislation is evolving to ensure that people have access to their land and that their human rights are protected. The discovery of new things on an economically important development tends to be silenced for monetary considerations (Bowker & Star,

1999). We observe a link between the government's social practices and data gathering with Dataset B and its category placement, which reveals some of these monetary reasons that represent the data's institutionalized silence. The interests of the government are obscured within the placement of data under the "national mapping" category. "It becomes clear that the technical infrastructure combined with race-ethnically categorized data are inherently racing systems, regardless of their intended purposes" (Van Schie, 2018). The need for the government to map Indigenous territories goes beyond the protection of Indigenous rights as the government claims. The overshadowing of data acquired on this dataset and its obfuscated placement under this category correspond to the government's history of corruption. It also runs counter to one of the digital government's most crucial goals: *transparency.*

There are ongoing examples of this dataset that demonstrate how Indigenous communities' government discrimination is evident in the setting of the Colombian government's data platform. Dataset B includes details about culture, history, and place. Then, as previously stated, "what are we mapping"? What makes you think this dataset belongs in the national mapping category? On one hand, at a "transparent level", the government is mapping Indigenous communities whose territories are internationally protected. However, at a "hidden" level, I found that the government is mapping its interest regarding the exploitation of gold while trying to cover important information with other irrelevant data. This will be explained in the next section.

Culture

Figures 1.10 and 1.11 show how I mapped the geo-reference of Dataset B's Indigenous communities. It was clear that the population was concentrated in three main villages: San Lorenzo, Riosucio, and El Salado. The collection clearly shows indigenous

geographic areas, but it also conceals critical information that has previously been covered by the government.
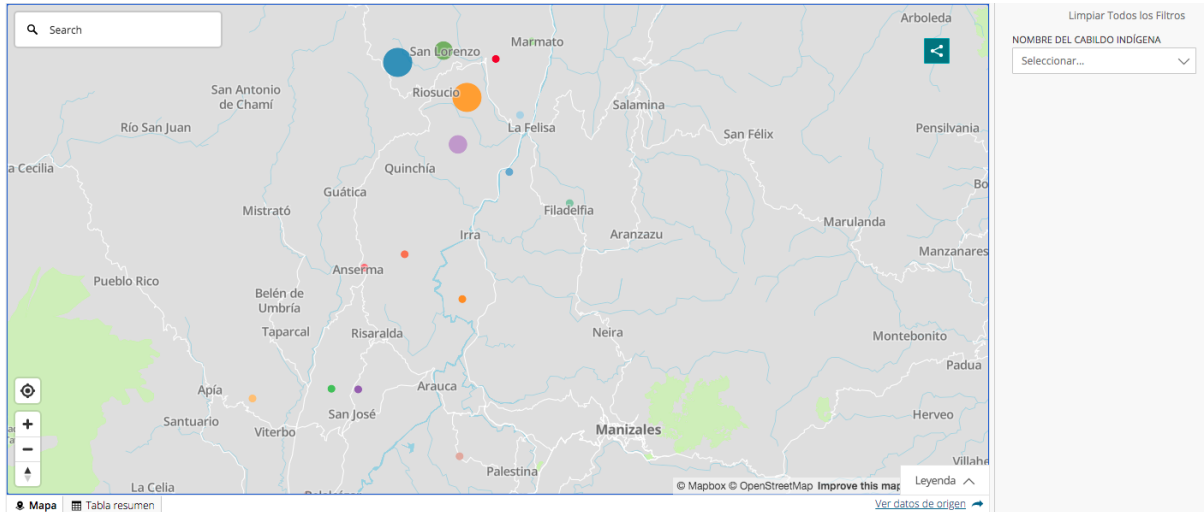
Figure 1.10 Geo-reference population

Figure 1.11 Geo-reference population - populations density

"Redacción el Tiempo" (2002), Colombia's biggest newspaper since the 1980s, has reported various atrocities committed to these specific communities in such territories. In 2001, a massacre against the leaders of the community killed seven Indigenous persons and a minor disappeared. In the same year, "En san Lorenzo" (2018) reported that illegal armed groups caused 135 individuals to flee their homes. Two years later, 150 guerrilla fighters

descended on the town, demolishing the educational facility and murdering a 14-year-old girl. The ex-guerrilla members' continued aggression, along with the lack of official protection, has resulted in a series of atrocities and criminality. As if that weren't enough, "Colombia en Transición" (2020) revealed that the Special Jurisdiction for Peace (JEP) declared that there could be more than **100 bodies** of those who went missing during the conflict on this land.

After researching historical events and a variety of datasets, it's clear that the Colombian government's discrimination isn't isolated within a single category, but rather can be found throughout the open data platform in numerous datasets, data gathering techniques, and categories. Just as Van Schie, in this research, "we argue for the use of critical frameworks through which the embedded normative assumptions of datafied systems and epistemologies can be made visible" (Van Schie, 2018). "Social Inclusion and Reconciliation" portrayed how Colombia's government includes data of Indigenous communities to give tangible results to international organizations. We find a series of trends in colonial discourse after evaluating the second category. It appears that by including culture in this dataset, there is a need to disseminate information that is useful to the government. In an imposed, purified system of categories, this "hidden data" carries many ironies and much individual suffering (Bowker and Star, 1999). The goal of "National Mapping" is to map more than simply geographic facts in order to enhance people's lives. It's a map of victims, erroneous government decisions, and a series of crimes that have yet to be prosecuted. When 'inclusive' information such as Indigenous rituals, dances, and sports is added to the dataset, it obscures certain other essential information about their current needs. When data treats Indigenous people as a "group" rather than individual citizens with rights, the categorization system shifts from inclusive to race-ethnically discriminatory. (Van Schie, 2018). The biggest concern with this is that "to increase levels of trust, the whole data ecosystem will need to build ethical considerations into how data is collected, managed, and used to ensure equity

around [...] how the benefits are distributed" (Dodds and Wells, 2019). However, I highlight the government's carelessness towards Indigenous Communities through historical facts, resource exploitation, and categorization ideologies, making "transparency" a tough goal to achieve.

THE UNSEEN OF MY TERRITORY

The Colombian government's data platform may have a greater harmful influence on their lives than it can improve by repeating discriminatory activities. With the proposed research, it is possible to engage with the technicalities that reproduce inequality, but also to comprehend how these affect our knowledge about colonial history and the post-war era (Van Schie, 2018). Colonialism is still alive and well in Colombia, whether through legislation or power dynamics. Indigenous peoples see their homelands as a place rich in history, ancestors, and a community that shares and belongs together. The government, on the other hand, gains from the capacity to create laws in order to exploit these territories. They exercise sovereignty and use this power to cover up government corruption. "Coloniality in this sense is not merely tied to a particular time or spatial territory but should be understood as a psychological, political, economic, social, epistemic, or ontological condition" (Van Schie et al., 2020). The datasets match the categories and depict Indigenous populations' discourses, such as being blamed for the country's inability to "advance," or how they hinder foreign investment in mining and exploitation. Latin American indigenous peoples face numerous challenges and are frequently powerless. They still protest the exploitation, illegal appropriation, and other forms of institutionalized discrimination (Jackson, 2007). Open data practices align with ongoing power relations and they function as a way to create social constructs independently of their principal established functions (Van Schie et al., 2020). Open data could become a tool to institutionalize power and how Indigenous are perceived in

society. Indigenous communities need repair from the war and conflict and the government has failed to recognize, accept, and repair the victims. The government also fails to use its data platform to repair these communities by relocating datasets into categories that do not correlate to their needs. They need a "shift in argument" from "rights as minorities" discourse to one claiming "rights as peoples" (Jackson, 2007). Indigenous argue that from a position that claims inherent rights, which derive from their status as autochthonous peoples, the government uses discourses that assimilate such populations with minority rights (Jackson, 2007). As a result, the government uses a data platform to provide various sorts of discrimination, although these prejudices go unreported by these populations, which are typically labeled minorities.

The disappeared bodies that lay under Indigenous territory are an analogy with the lack of data transparency from the government. The unseen bodies are a metaphor for the unseen intentions of the data collection and categorization created by the government. The lack of transparency in the data is similar to the lack of transparency of the 'legitimate' intentions of the government with the Indigenous territories. They utilized Indigenous protected areas to cover up the atrocities of the government-supported war, much as they appear to use 'cultural data' to cover up the reasons for "mapping" Indigenous territories. Indigenous peoples were oblivious of the concealed bodies beneath their sacred ground, much as they are uninformed of the impact of digital infrastructure on their life. "This means that from that moment onward, data and information were prioritized over the bodies of the people to which the data referred" (Van Schie et al., 2020). During my investigation, I discovered that, while there were five suitable datasets in each of these two categories, there were more than five **uncategorized** datasets related to Indigenous issues. It will be challenging to maximize the social and economic value of this data platform without trust

(Dodd and Wells, 2019). These behaviors will go unnoticed unless the government intervenes to help the true victims of the conflict.

CONCLUSION

With international support and financial aid, Colombia's peace agreement created a series of initiatives and social plans to develop the country in a post-conflict era. The Colombian administration realizes that corruption and war have created a climate of distrust among Colombians. The government seeks to lessen this by creating a *digital government* that provides free access to public data. They encourage the research and investigation of the data available to monitor the transparency of the public sector. The proposed research evaluated two categories of open data. In the category "Social Inclusion and Reconciliation" we found that the data collection system aligns with the disparity of datasets under the category. This creates a misplacement of "communities' needs" under a category that prioritizes other social matters. The Colombian government's data platform reflects social problems such as emphasizing foreign affairs above domestic ones, developing services that are insufficient for Indigenous living situations, and monetizing Indigenous "international protection" and recognition. The data system depicted in Dataset B is analogous to the state's territorial and political dispute with Indigenous people, according to the category "National Mapping". More crucially, the government's interests are hidden within the data system, contradicting the data platform's goal of providing transparency to Colombian residents. The categorization of datasets exposes socio-political ideologies of ongoing colonialism that promote inequality and accentuate power relations.

For future research, it is important to further understand the categorization within the datasets and what they represent for these communities. There are more uncategorized Indigenous community datasets in open data than those that fall into a category. We

advocated researching uncategorized datasets in order to have a better understanding of their discursive practices. We propose to the Ministry of Information Technologies and Communications (MinTIC) to review the classification of these datasets and contemplate the possibility to allow the platform to place datasets under multiple categories at once. This allows the inclusion of all sectors that should be involved with specific sets of information.

Regarding the impacts found in this research on Indigenous communities, it is difficult to suggest one solution. The ambivalence of data could improve communities' lives but could also allow for power relations to strengthen. In the context of the Colombian government's data platform, there may not be a single right response for addressing the effects of political discrimination against Indigenous populations. It may be possible to monitor more data from Indigenous communities and provide services to improve their lives if data systems become more consistent and fair across the country. However, it may open up more opportunity for ideologies to continue reproducing, resulting in more inequality and prejudice. Although we encourage the government to establish data collection processes together with citizens' participation, it is important to understand each community's needs separately. Although there is no guarantee that every community will want to participate in data collection systems, future ethnographic study will allow us to better understand the amount of participation that the government needs to attain. This will also allow us to understand how to promote the good practices of open data while better understanding to what extent Indigenous should become actors of the data ecosystem.

REFERENCE LIST

- Anderson-Smith, B. (2008). Exploring engagement between Indigenous communities and government: lessons for country management. The Australian National University. Retrieved from: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.461.7459&rep=rep1&type= pdf.

- Beteille, A. (1998). The idea of indigenous people. *Current Anthropology, 39,2*. Doi: 10.1086/204717

- Beitte, J. S. (2015). Peace talks in Colombia. Congressional Research Service. Retrieved from https://fas.org/sgp/crs/row/R42982.pdf

- Bowker, G. C., & Leigh, Susan. (1999). Sorting Things Out: Classification and Its Consequences. MIT Press. Retrieved from https://ar264sweeney.files.wordpress.com/2015/09/9780262269070_introduction.pdf

- Colombia en transición. (2020, May 13). Resguardo indigena de San Lorenzo (Caldas): el nuevo lugar donde la JEP busca desaparecidos. *El Espectador.* Retrieved from https://www.elespectador.com/

- Colombia reconoce que más de 20.000 víctimas fueron víctimas del conflicto armado. [Editorial]. (2020). *DW News.*

- Constain Rengifo, S., Rueda Jimenez, G., Patricia Pico, C., Media, L. F., Escobar, A. C., Matamoros, C., Santofimio, N., Vargas Tamayo, A., Rodriguez, C, M., Aponte, J. R., & Leon Caideo, C. J. (2019). Guía para el uso y aprovechamiento de Datos Abiertos en Colombia. *Gobierno de Colombia MinTIC.* Retrieved from: Guia de Datos Abiertos de Colombia.pdf

- Constain Rengifo, S., Castro Sierra, JBolanos, C., RNoriega Silva, J. C., Rodriguez Alvarez, C., Santofimio, N., Parada Gallardo, J. C., Vargas Tamayo, A. (2018).

Manual de gobierno digital: Implementación de la política de gobierno digital. *Gobierno de Colombia MinTIC*. Retrieved from: Guia de Datos Abiertos de Colombia.pdf

- Constain, S. C., Sierra, J. C., Bolanos, C. R., Noriega, J. C., Rodriguez, C., Santofimio, N., Parada, J. C., Tamayo, A. V. (2018). Manual del gobierno digital. *Ministerio de Tecnologías de la información y las comunicaciones*. Retrieved from Act 1078 of 2015, chapter 1, title 9, Part 2, book 2.

- Daniel Lathrop and Laurel Ruma, Open Government: Collaboration, Transparency, and Participation in Practice (Sebastopol, CA: O'Reilly Media, Inc., 2010); Tony E. Wohlers and Lynne Louise Bernier, Setting Sail Into the Age of Digital Local Government: Trends and Best Practices (New York: Springer, 2015).

- Davies, T., Walker, S., Rubinstein, M., & Perini, F. (Eds.). (2019). Data Infrastructure. The State of Open Data: Histories and Horizons. Cape Town and Ottawa: African Minds and International Development Research Centre.

- Dawes, S. S., Vidiasova, L., & Parkhimovich, O. (2016). Planning and designing open government data programs: An ecosystem approach. *Government Information Quarterly*. 33 (1), 12-2. Doi:10.1016/j.giq.2016.01.003.

- Dodds, L & Wells, P. (2019). Data infrastructure. Davies, T., Walker, S., Rubinstein, M., & Perini, F. (Eds.), The State of Open Data: Histories and Horizons (pp 260-273). Cape Town and Ottawa: African Minds and International Development Research Centre.

- En San Lorenzo (Riosucio) con empuje entierran la guerra. (2018, May 9). La patria.com. Retrieved from https://www.lapatria.com

- Fairclough N (1995). Critical Discourse Analysis. The Critical Study of Language. London: Longman

- Gray, J., Gerlitz, C., & Bounegru, L. (2018). Data Infrastructure Literacy. *Big Data & Society, 1-13*. Doi:10.1177/2053951718786316.

- Guasca, M., Vanneste, D., & Van Broeck, A. M. (2021). Peacebuilding and post-conflict tourism: addressing structural violence in Colombia. *Journal of Sustainable Tourism*. DOI: 10.1080/09669582.2020.1869242

- Histrov, J. (2010). Indigenous Struggles for Land and Culture in Cauca, Colombia. *The Journal of Peasant Studies*, 32:1, 88-117, DOI: 10.1080/0306615042000322402

- Idrobo, N., Mejia, D., & Tribin, A. M. (2014). Illegal gold mining and violence in Colombia. Peace Economics, Peace Science and Public Policy, 20 (1). Doi: 10.1515/peps-2013-0053

- Jackson, J, E. (2007). Rights to Indigenous Culture in Colombia. In M.G and S. E. M (Ed.), *the practice of human rights: tracking law between the global and the local (pp.204-241)*. Cambridge20: Cambridge University Press.

- Janks, H. (1997). Critical discourse analysis as a research tool. Discourse: Studies in the Cultural Politics of Education. 18(3), 329-342.

- Jorgensen, M., & Phillips, L. J. (2002). *Discourse Analysis as Theory and Method.* London: Sage Publications.

- Kitchin, Rob (2014). The Data Revolution: Big data, open data, data infrastructures & their consequences. London: *Sage*. p. 49. ISBN 978-1-4462-8748-4.

- Mauro, F., & Hardison, P. D. (2000). Traditional knowledge of indigenous and local communities: international debate and policy initiatives. *Ecological Applications, 10, (5) 1263-1269.* doi:10.1890/1051-0761(2000)010[1263:TKOIAL2.0.CO;2

- OECD (2017), Assessing the Impact of Digital Government in Colombia: Towards a new methodology, OECD Digital Government Studies, OECD Publishing, Paris, https://doi.org/10.1787/9789264283282-en.

- Open Definition. 2019. "The Open Definition - Open Definition - Defining Open in Open Data, Open Content, and Open Knowledge." https://opendefinitio n.org/.

- Phelan, S. (2017). critical discourse analysis and media studies. F. J and R. J. E. (Ed), The Routledge handbook of critical discourse studies (285-290). Oxon: Routledge.

- Radway, R. J. (1980). Transfer of technology to Colombia: A proposal to modify decision 24. *Lawyer Of The Americas, 12, 321.* Retrieved from: https://heinonline-org.proxy.library.uu.nl/HOL/Page?handle=hein.journals/unmialr12 &id=1&collection=journals&index=

- Ramanathan, R., & Tan, B.H. (2015). Application of critical discourse analysis in media discourse studies. 3L: The Southeast Asian Journal of English Language Studies – Vol 21(2): 57 – 68. Retrieved from: https://ejournal.ukm.my/3l/article/view/9183

- Redacción de El Tiempo. (2001, april 19). Exterminio de Indígenas. El Tiempo. Retrieved from https://www.eltiempo.com/archivo/documento/MAM-1373459

- Rettberg, A. (2019). Peace-making amidst an unfinished social contract: The case of Colombia. *Journal of Intervention and Statebuilding, 14 (1, 84-100).* Doi: 10.1080/17502977.2019.1619655

- Richard Rottenburg and Sally Engle Merry, "A World of Indicators: The Making of Governmental Knowledge through Quantification," in The World of Indicators, ed. Richard Rottenburg et al. (Cambridge: Cambridge University Press, 2015), 15.

- Roldan Ortega, R. (2004). Models for recognizing Indigenous land rights in Latin America. The World Bank Environment Department, 99. Retrieved from: https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.849.903&rep=rep1&type= pdf

- Rodríguez-Garavito, C. A., & Arenas, L. C. (2005). Indigenous rights, transnational activism, and legal mobilization: the struggle of the U'WA people in Colombia. In S. Boaventura, & C. Rodriguez (Eds.), *Law and Globalization from Below: Towards cosmopolitan legality (pp.241 - 395).* Cambridge University Press.

- Ruppert, E (2015) Doing the transparent state: Open government data as performance indicators. In: Rottenburg, R, Merry, SE, Park, SJ(eds) A World of Indicators: The Making of Governmental Knowledge through Quantification, Cambridge: Cambridge University Press, pp. 1–18.

- Star, S. L., & Ruhleder, K. (1996). Steps toward an ecology of infrastructure: Design and access for large information spaces. *Information System Research,2 (15).* Doi: 10.1287/isre.7.1.111

- Schrank, S. (2004). Golden Gate Bridge. *The Journal of American History; Oxford.* 91 (3) 1132-113. Retrieved from https://www.proquest.com/pubidlinkhandler/sng/pubtitle/The+Journal+of+American+History/$N/41964/OpenView/224898385/$B/B6E4E0E23A764766PQ/1;jsessionid=0D70F9D704C73B4C544A1157D11087AB.i-0be77e120ae1a743f

- Trindade, M. (2008). International Cooperation and Internal displacement in Colombia: facing the challenges of the largest humanitarian crisis in South America.

- 2014b. "Big Data, New Epistemologies, and Paradigm Shifts." Big Data & Society 1 (1): 205395171452848. https://doi.org/10.1177/205395171 4528481.

- Van Schie, G. (2018). Origins: A History of Race-Ethnic Categorisation in the Dutch Governmental Data Ontology (1899 - 2018). *Tijdshrift voor mediageschiendenis, 21 (2).*

- Van Schie, G., Smit, A., & Coombs, N. (2020). Communication and Media Racing through the Dutch Governmental Data Assemblage: A Postcolonial Data Studies Approach. University of California Press, 1 (1). Doi:10.1525/gp.2020.12779.

- Wiessner, S. (2011). The cultural rights of indigenous peoples: achievements and continuing challenges. European Journal of International Law, 22 (1), 121-140. Doi:10.1093/ejil/chr007

- Zuiderwijk, A., & Janssen, M. (2014). Open data policies, their implementation, and impact: A framework for comparison. *Government Information Quarterly, 31, 17-29.* Doi: 10.1016