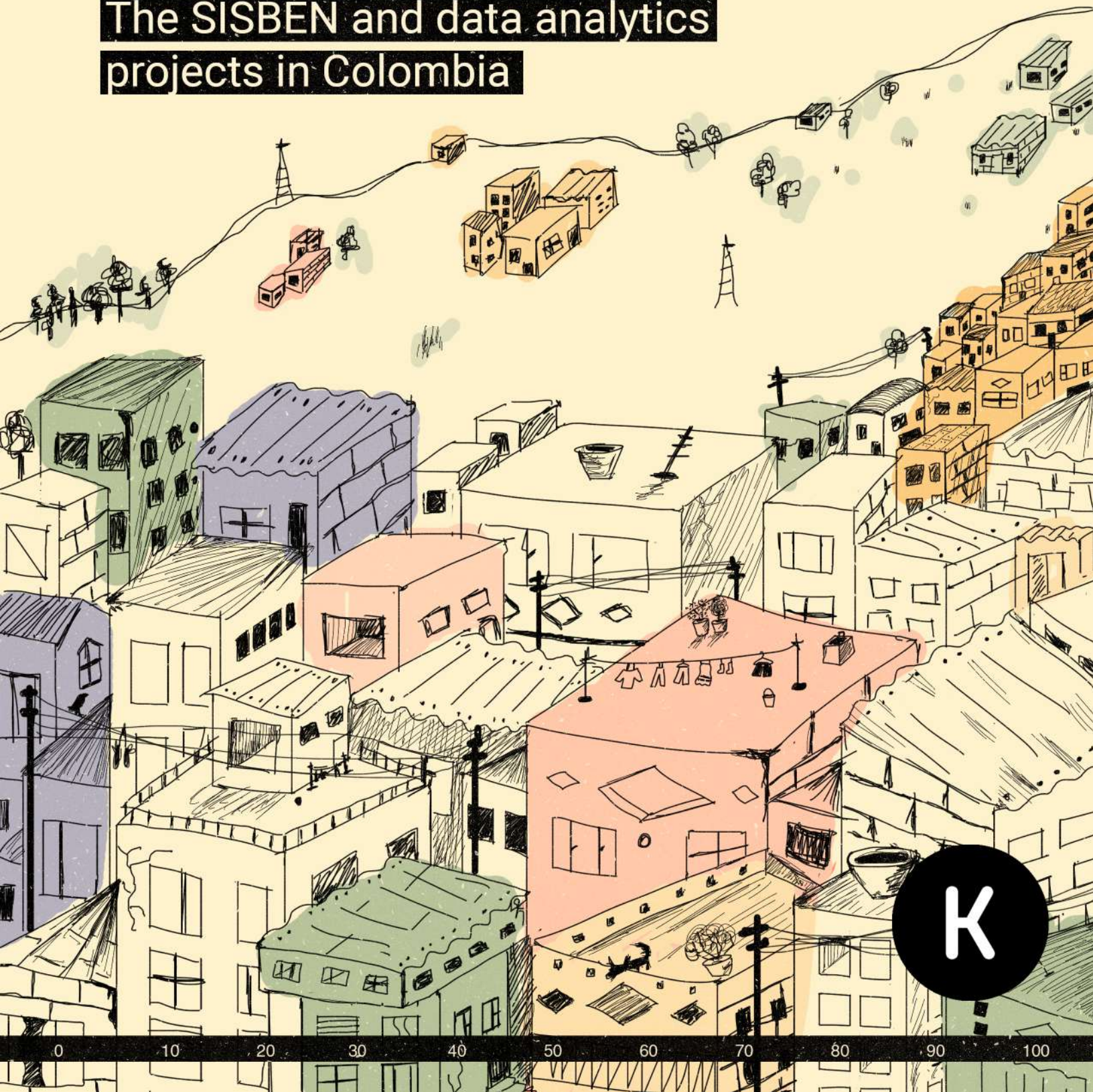


EXPERIMENTING WITH POVERTY:

The SISBEN and data analytics projects in Colombia



EXPERIMENTING WITH POVERTY:

The SISBEN and data analytics projects in Colombia

Author: Joan Lopez

Coordination: Juan Diego Castañeda

Fundación
Karisma

Bogota, Colombia 2020

In an effort so that all people have access to knowledge, Fundación Karisma is working to make its documents accessible. Its format, therefore, includes metadata and other elements that make it compatible with tools such as screen readers or braille displays. The purpose of accessible design is that everyone can read, including those who have some type of visual impairment or difficulty in reading and comprehension.

More information on accessible documents at:
<http://www.documentoaccesible.com/#que-es>



This material circulates under a CC BY-SA Creative Commons license

4.0. You may remix, retouch, and build upon the material for any purpose, even commercially, as long as you credit the author and license the new creations under the same conditions.

To view a copy of this license, visit: https://creativecommons.org/licenses/by-sa/4.0/deed.es_

EXECUTIVE SUMMARY

The SISBEN is an instrument for managing the population living in poverty that uses a scale from 0 to 100 to rate its beneficiaries in terms of prosperity: 100 is “more prosperous” and 0 “less prosperous.”

The information system used by this system to calculate each person’s score is collected through a survey in the areas where poverty has traditionally been located and is the input for each of the entities that administer one of the 18 State social benefit programs, determines if a person can apply for access to one or more of them.

In 2016 the new version of SISBEN was created, and with this, substantial changes gave an account of the interest that the current government has in experimenting with the data and information of people living in poverty.

On the one hand, the National Planning Department (DNP), the entity that administers the SISBEN database, decided to modify this system’s algorithm to include the prediction of “income-generating capacity.” An attempt to reduce the number of people living in poverty who could access a social benefit was made this way. On the other hand, an information exchange system was created with 34 public and private databases to verify citizens’ reported data.

Additionally, the survey was changed so that the refusal to supply information by people living in poverty would result in their exclusion from the system and that any inconsistency in the information could lead to legal and administrative actions.

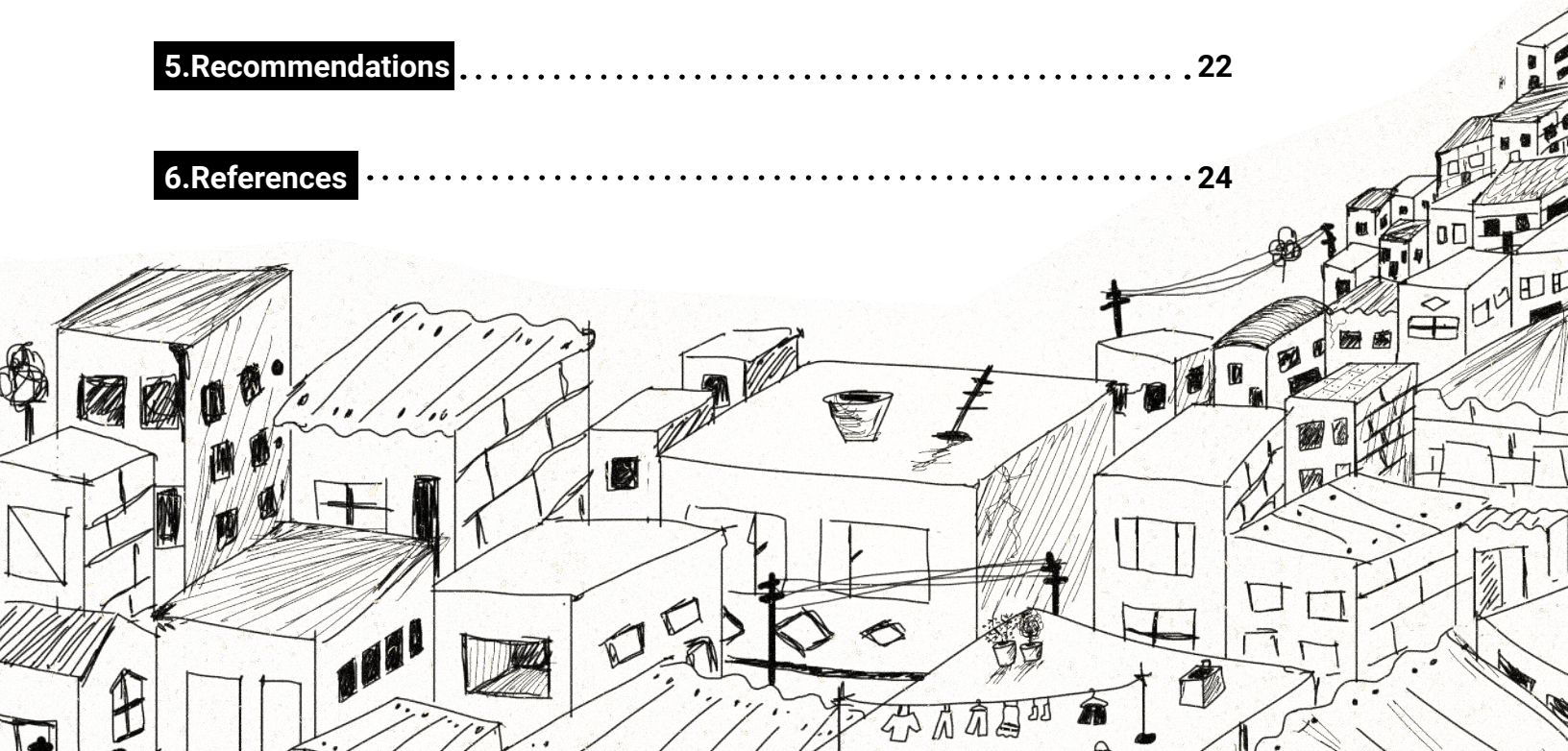
In the text that follows below, we also analyzed two cases that show the changes in logic and the experimentation processes with the SISBEN data.

The first was an agreement with the Irish multinational Experian to use databases from the financial risk rating agency Datacredito, establishing that the former would allow and facilitate DNP access to their Quanto solution – with which a person’s income levels can be estimated. Consent was also negotiated for viewing financial information of the holders who are registered in the DataCrédito databases to search



CONTENT:

1. Introduction	5
2. The System of Possible Beneficiaries of Social Programs (SISBEN)	7
a. The SISBEN as a social classification algorithm	7
b. SISBEN and resource targeting	8
c. Why is the SISBEN score important?	8
d. The fourth version of the SISBEN and its changes	10
3. New technologies and experimentation with social policy data	16
a. The Experian case and financial profiling	16
b. The INNPULSA Colombia case: Big Data and innovation with social policy	18
4. Conclusions	20
5. Recommendations	22
6. References	24



1. INTRODUCTION

Activism on technology and digital rights issues is experiencing a moment of transformation, which is about building bridges with other political agendas (Dencik, Hintz, & Cable, 2016). However, it has been difficult to establish the relationship between the struggle between human rights and social inequalities within the framework of social justice (Rodríguez Garavito, 2019). A vision of any technology considered from the global south must bring a concern for social inequities and violence towards certain groups by the State (Arora, 2019a; Milan & Treré, 2019). For this reason, several authors have constructed the concept of “data justice” to refer to the connection between the technical possibilities of digital technologies and the social justice agenda understood as the fight for a less inequitable society that protects the social, civil, and political rights of the people (Dencik et al., 2016; Dencik, Hintz, Redden, & Treré, 2019; Heeks & Renken, 2018; Taylor, 2017).

A problem that can be read from this perspective is the development, possibilities, and limits of Artificial Intelligence. As always, there are enthusiasts who consider that societies could solve all kinds of problems using these technologies. Likewise, there are voices that consider that its use is a risk for the exercise of rights such as privacy and non-discrimination. However, the problem seems distant from our realities, far from the latest technologies controlling our lives in a dystopian future. Without a reflection that connects our contexts with those of countries of the global north, discussions could get caught up in speculation over future risks or in defining, for example, if the technology that governments want to apply is a “true” Artificial Intelligence (AI).

When we think about social justice within data systems, one begins to notice that abstract AI questions leave out classification and discrimination practices through states’ data in the global south. Despite not being sophisticated technologies, these systems involve an effort to automate people’s scoring, predicting behaviors, and profiles that determine access to essential public or private goods and services for people. The lack of “real” implementation of Artificial Intelligence solutions does not mean that it is unnecessary to track the initiatives that promise to use them. On the contrary, it is essential to continue with the documentation work and continue questioning the authorities from these plans’ announcement until their implementation.

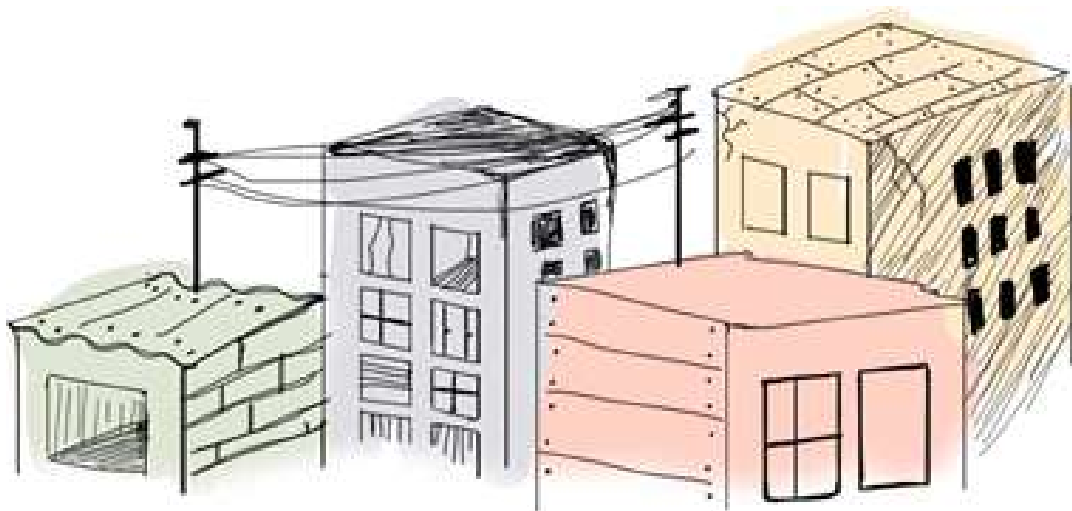


In Colombia, a system collects personal data, classifies people according to their level of poverty on a numerical scale, and uses this score to decide if they deserve access to state benefits and services. The SISBEN, which is the main instrument for

allocating social benefits in Colombia and has existed for more than two decades, has not been analyzed from a social justice perspective in data systems. From this paradigm, we see SISBEN as an intensive system in using personal data that has left us with manifold questions about how this data is used to monitor people and open or close the doors to state benefits.

This document offers an analysis using multiple sources, including academic articles, petition rights, public policy documents, contracts and state agreements, SISBEN training materials, news, press releases, and base figures from public data. The text consists of three parts. First, we make (1) a description of the SISBEN, and we stop at the latest modifications so that the system uses more data to profile people. Next (2), we explain two projects leveraging beneficiaries' data that potentially go against the data protection law and constitute examples of social injustice in managing information from State programs. Finally (3), we note some conclusions and recommendations.

We now highlight the need to connect the global concerns of Artificial Intelligence with practices in the global south that, despite not having sophisticated technologies, involve scoring people, predicting behaviors that affect participation in social life, and respect for human dignity.



2. THE SYSTEM OF POSSIBLE SOCIAL PROGRAM BENEFICIARIES (SISBEN)

a. The SISBEN as a social classification algorithm

The SISBEN is an instrument for classifying the population based on its socioeconomic achievements that was first used in 1994 (Castañeda & Fernández, 2005; Sarmiento et al., 1999; Vélez, Elkin Castaño, & Deutsch, 1999). The system has two components: the individual collection of socioeconomic data of people and the classification of the said population from 0 to 100, where 100 is “more prosperous” and 0 “less prosperous” (Sarmiento et al., 1999). The resulting score will be used by each entity that administers social benefits, such as the Ministry of Health, the Ministry of Labor, or Social Prosperity, to determine if a person can request access to a social program.

The information that SISBEN uses to calculate each person’s score is collected employing a survey. The survey is applied to people located in places considered that there may be a concentration of low-income groups from the analysis of socioeconomic information (Vélez et al., 1999). The people who are considered vulnerable and do not reside in these areas can ask to be surveyed. After the information is collected, specialized software is used that generates the individual score and the ordering of the population (Castañeda & Fernández, 2005; Sarmiento et al., 1999).

The National Planning Department (DNP) is an entity that reports directly to the Office of the President and is the administrator of the SISBEN database. In addition, it is in charge of updating the algorithm that generates the individual scores. To do this, it must determine the values of each category and the elements that are measured (Castañeda & Fernández, 2005). The first three designs of the SISBEN measured a set of characteristics of the population that include demographic variables, consumption of durable goods, human capital and current income (Vélez et al., 1999). These components were based on a vision of poverty focused on living standards (Menjura Murcia, 2016).



b. SISBEN and resource targeting

The SISBEN appeared in 1994 as a product of various reforms during the 90s, which were driven by the debt crisis in Latin America (Sarmiento, González, & Rodríguez, 1999). In this way, in the midst of the need to receive international economic support, governments began a rapid process of liberalization of the economy and social policy (López Restrepo, 1995) in which the positions promoted by the World Bank and the International Monetary Fund were elementary for the reforms that were made that decade in the region (Carnes & Mares, 2015; Deacon, 2007).

In the 1990s, the World Bank began to promote the targeting of resources, the privatization of the provision of public services, and the austerity policies of public spending (Hall, 2007). For this reason, the idea that the State began to “look for poor people” became a fundamental concern to focus the limited resources left from tax and tariff reforms (McGee, 1999). With this, the State began to require more data to “find” poor people and especially to target resources individually.

As universal social benefits or those aimed at all citizens were abandoned, a welfare state was preferred that only supports the most impoverished segments of the population (Carnes & Mares, 2015; Deacon, 2007). Thus, the SISBEN began to take on greater importance and became the main instrument for targeting public resources.

c. Why is the SISBEN score important?

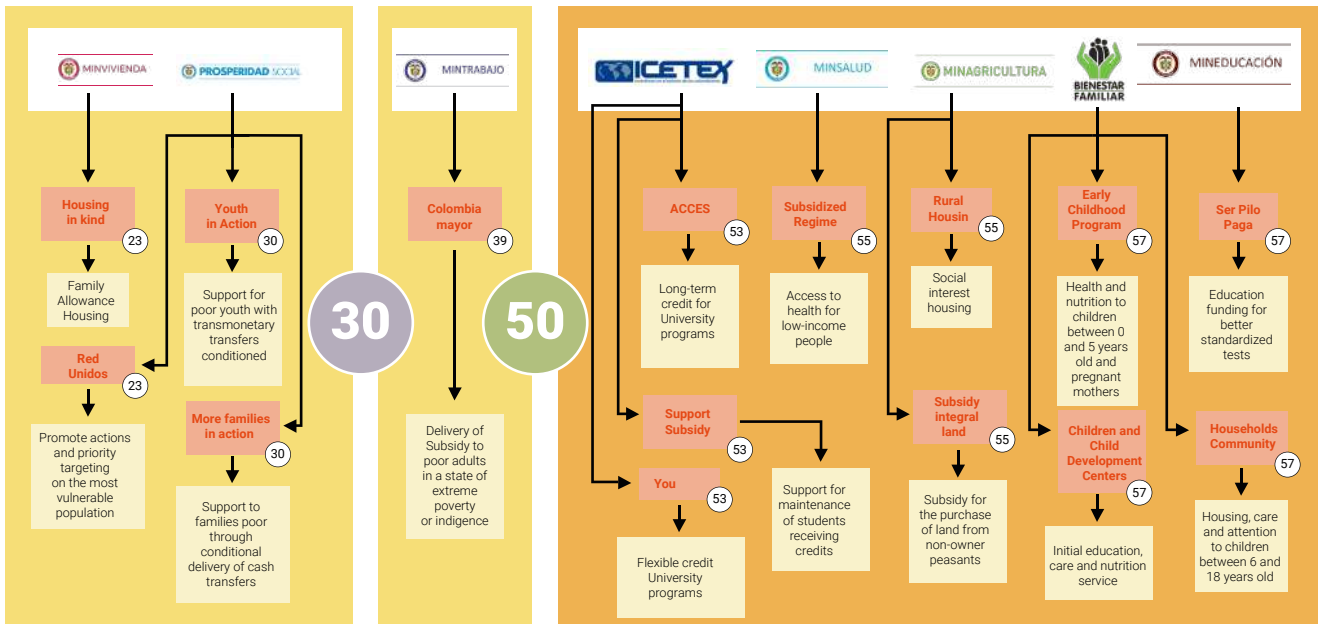
As said, the SISBEN is an individual scoring system that determines who may be “worthy” of the social benefits of the Colombian State. Since it was created, the list of benefits administered nationally has not stopped growing and the territorial entities increasingly depend on SISBEN for their own programs. Currently, at the national level, it is used for 18 social programs of different characteristics (See Diagram 1). Most benefits require processes for selection and establishment of program-specific requirements such as cut-off points, special characteristics of the population and budget availability. In that sense, many people access more than one program (CONPES, 2016) and the stability of different aspects of their lives such as health, education, old age, housing and income depend on the SISBEN score.

One goal of the government is to abandon socioeconomic classifications of strata of real estate by zone and to rely entirely on individualized scores such as the SISBEN to deliver all social rights. An example is the “Social Registry” project that is present in the National Development Plan (2018-2022) of the current government and that will try to include more data in the SISBEN profile, such as some benefits that currently do not include individualized targeting such as public education and public utilities subsidies (National Planning Department, 2019a; Morales Manchego & Galindo Caballero, 2019).



Diagram 1:

Most important programs with a defined cut-off point (CONPES, 2016)



As said, each entity decides how much is the necessary score to request a social benefit. Despite the fact that, for the DNP, below 50 points is considered equivalent to being in poverty situation, only five programs (Housing in Kind, United Network, Families in Acción, Jóvenes en Acción and Colombia Mayor) require a score in this range. When analyzed in detail, four programs (Housing in Kind, United Network, Families in Action, Youth in Action) focused on poverty alleviation, are below 30 points and one is below 39 (Colombia Mayor). Likewise, programs above 50 points, or, they seek to be universalists¹ such as the subsidized health regime and communal households or they try to be very focused due to the reduced amount of benefits such as Ser Pilo Paga and credits for higher education (See Diagram 1). In that sense, the programs above 39 points are centered on another type of population that is not necessarily in a situation of poverty.

The SISBEN turns out to be a key classification for the lives of millions of people in Colombia that basic services cannot be provided and depend on state aid. For this reason, your refusal to provide all your data, or that there may be errors in them, would imply serious consequences for people living in poverty. This situation is quite worrying for a country like Colombia, with 10.8% of the population unem

1. The idea of these types of services is that they reach all the people who are not in a position to pay or contribute to the services. For example, the subsidized health regime has a logic in which health is a fundamental right (Law 1751 of 2015) and aims for all people to have coverage and those who cannot contribute to the system must be subsidized by the State.

ployed and 47% in informal jobs (DANE, 2019a, 2019b). This situation has led many people to refuse to be interviewed out of fear their score changes (Conexion Sur, 2019; Diario del Cauca, 2018) and for DNP officials to threaten potential beneficiaries saying that “only be those who tell the truth will stay” (La Opinion, 2019).

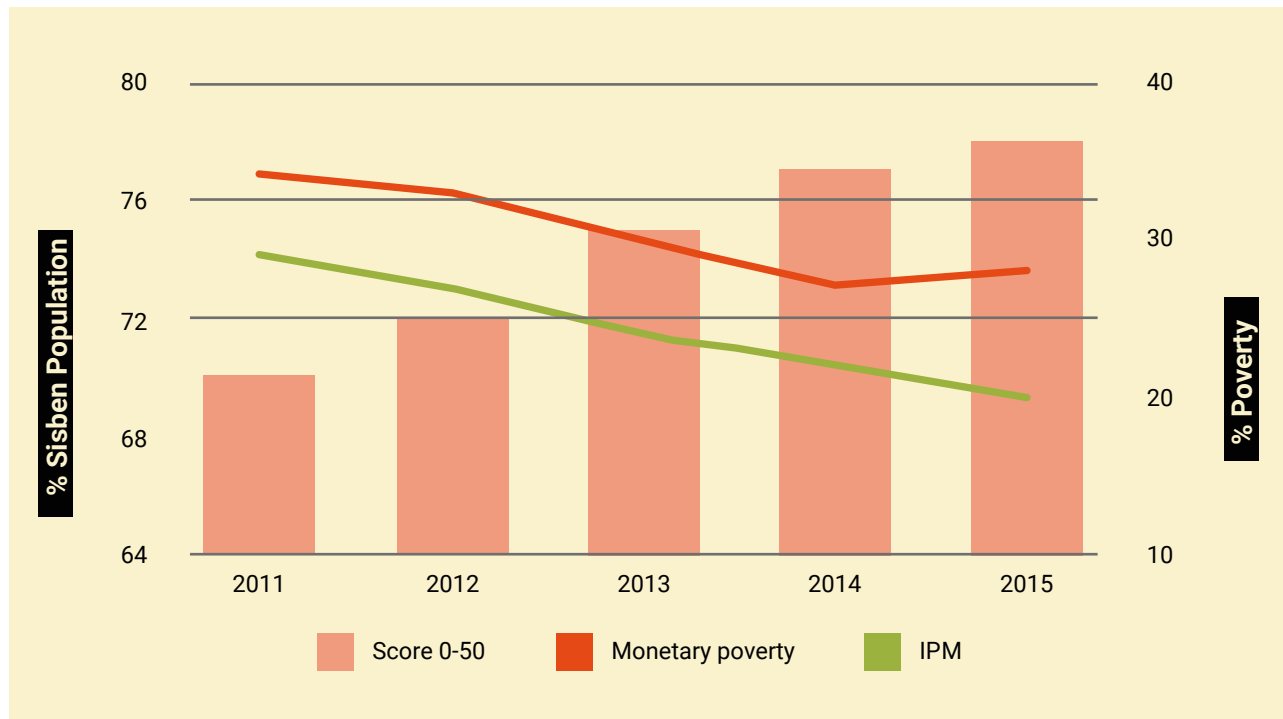
d. The fourth version of the SISBEN and its changes

The SISBEN and its methodology must be updated from time to time. So far, four versions of this system have been made since its creation in 1994. In 2016, for the formulation of the fourth version of the system, in an analysis carried out by the DNP with support from the World Bank, the Economic Commission for Latin America (ECLAC) and two external consultants, two basic problems were found with SISBEN: the lack of the component of “income-generating capacity” and the lack of an interoperable system to verify the information reported by citizens (National Planning Department, 2019b).

In the first place, according to the DNP, the SISBEN score as a standard of living is failing because, while the measurement of monetary and multidimensional poverty, based on representative samples has decreased, the percentage of the population with scores below 50 that would allow applying for some benefits has increased. The DNP considered it necessary to modify the algorithm with which the SISBEN score is performed to include the “income-generating capacity” and thus align the SISBEN score with the multidimensional poverty index (National Council for Economic Policy & National Planning Department, 2016). Figure 1 below shows how, according to the DNP, the percentage of people with a SISBEN score from 0 to 50 was lower in 2011 than that measured by the indices of monetary poverty and multidimensional poverty. As of 2014, according to this graph, there are more qualified people in the lower half of the SISBEN than people in poverty –monetary and multidimensional–.

Figure 1:

Incidences of monetary and multidimensional poverty and percentage of the population with a score low at the base of Sisben III (DNP, 2016)



The survey was modified to include questions about each person's economic activities, such as the amount and origin of income, expenses, and social benefits that are already received (National Planning Department, 2016). These new questions seek to provide a way to establish a profile of "presumption of income," which is, in essence, a prediction of people's income level. In the words of the DNP: "This methodology allows these two approaches to be taken into account in the calculation of the score, approximating the capacity of the population to generate income, with the purpose that the resulting classification responds to the behavior of monetary and multidimensional poverty of the country" (CONPES, 2016, p. 36). In other words, the SISBEN score now includes the probability that a person has a certain income level. This is different from verifying the income level or the living conditions of people.

The SISBEN algorithm acts as a "black box" in which we only know the input files and the results. Both the algorithm and the SISBEN's data treatment are unknown by the citizenry classified. To this end, we requested access to information from the DNP asking for greater clarity on prior studies that support the variables chosen for the SISBEN algorithm, the units of measure, and specificity of the variables taken into account to predict "income-generating capacity" (National Planning Department-DNP, 2019). However, the DNP responded that "the information is subject to confidentiality," since disclosing it may compromise "the macroeconomic and financial stability of the country" since "it may lead to modifying the information registered in the database constituting fraud" (National Planning Department-DNP, 2019, p. 4). While the State uses people's data for other purposes and, in turn, makes people's lives more transparent, the citizenry is less aware of how they are scored and the mechanisms with which the system works.

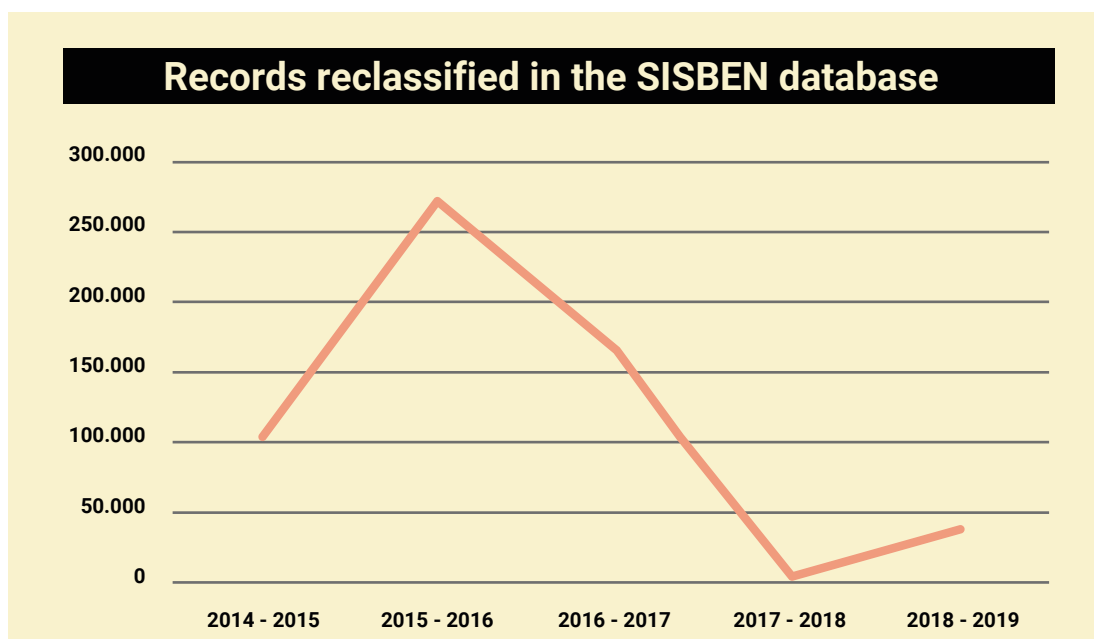
The second problem pointed out by the justification document for the latest version of SISBEN was the lack of an information exchange system to verify the citizen-reported data. For the DNP, this limits the state's ability to identify and avoid inconsistencies to make efficient use of public resources. Furthermore, entities cannot know the characteristics and benefits received by each person to make "possible the design of supply packages suitable for their needs" (CONPES 2016, p. 47).

To demonstrate the problem of user manipulation of the SISBEN, the DNP cross-checked the pension and health system database with that of the SISBEN. Through this effort, the entity marked 653 thousand cases "under verification" because they appear to show high income or because they appear as deceased persons (National Planning Department, 2019b). A public campaign followed to show the analysis results with cases of people with exorbitant salaries, deceased persons who were part of the system, and sudden changes in some people's homes (CONPES, 2016, p. 26). Far from looking like the State failing to manage its own databases, the government presented this effort as a planning victory thanks to technology against a group of malicious people in the system (Arbeláez Cama- year, 2019; Portfolio, 2017).



Figure 2:

Number of reclassified records in the database comparing annual modification
 Source: DNP, 2019.



To solve the problems of prior versions of SISBEN, CONPES (2016) established the need to create an information exchange system to carry out verification and validation processes of the information contained in the database of the SISBEN under the charge of DNP. The strategy, to this end, is to compare the records of different public and private entities to identify “possible inconsistencies” and allow the “automatic updating of the information” registered in the SISBEN database (p. 46). As reported by the DNP, the interoperability process is planned for execution with at least 34 databases that include categories such as Health, Pensions, Education, Work, Real Estate, Taxes, Financial risks, Social benefits, Transportation, Registry of Victims and Public Services (See Table 1).

Table 1:

Data from the SISBEN IV socioeconomic characterization sheet that could be verified with the new interoperability agreements (National Planning Department, 2019c).

Category in the SISBEN file	Type of data to be verified	Database for verification
Household identification	Geographic information	Agustín Codazzi Geographical Institute (IGAC)
Housing and household data	Natural disaster impact	IGAC
	Access to residential public utilities	Superintendency of services



Category in the SISBEN file	Type of data to be verified	Database for verification
Expenses	Copays and fees to the health system	ADRES (health system resource fund)
	Residential public utilities	Superintendency of services
	Bank loans and telecommunications services	Datacredito
	Pension contributions	Pension Management Unit (UGPP)
	Taxes and debts	Tax Department (DIAN)
Sociodemographic characterization	Names and surnames, date of birth, age, and sex	National Registry of Civil Status (RNEC)
	Victim of the conflict armed	Single Registry of Victims
	Civil Status and Relationship	RNEC
Health and fertility	Permanent health limitations	Individual Registry of Health Benefits (RIPS)
	Accidents, treatments, and medical care	RIPS
	Enrollment in the system health	ADRES/Enrollee Database (BDUA)
	Pregnancy	RIPS/Early Childhood Care
	Nutrition	Food Security Network
Education	Current educational status	Integrated Enrollment System (SIMAT) / National Higher Education Information System (SNIES) / National Apprenticeship Service (SENA)
	Highest educational level	SIMAT / SNIES / SENA
	Pension contribution	BEPS/UGPP
Occupation	Main economic activity	Integrated Contribution Settlement Worksheet (PILA) / UGPP / BEPS
	Job search	Public Employment Service
Income	Monthly earnings	PILA/UGPP/BEPS/ Chamber of Commerce (CC)/ Datacredito
	Monthly earnings	PILA/UGPP/BEPS/DIAN/ CC/ Datacredito
	Earning ratios	PILA/UGPP/BEPS/DIAN/ CC/ Datacredito
	Subsidies received	Housing/Families in Action / Youth in Action / Senior Colombia / Rural Subsidies / Ser Pilo Paga

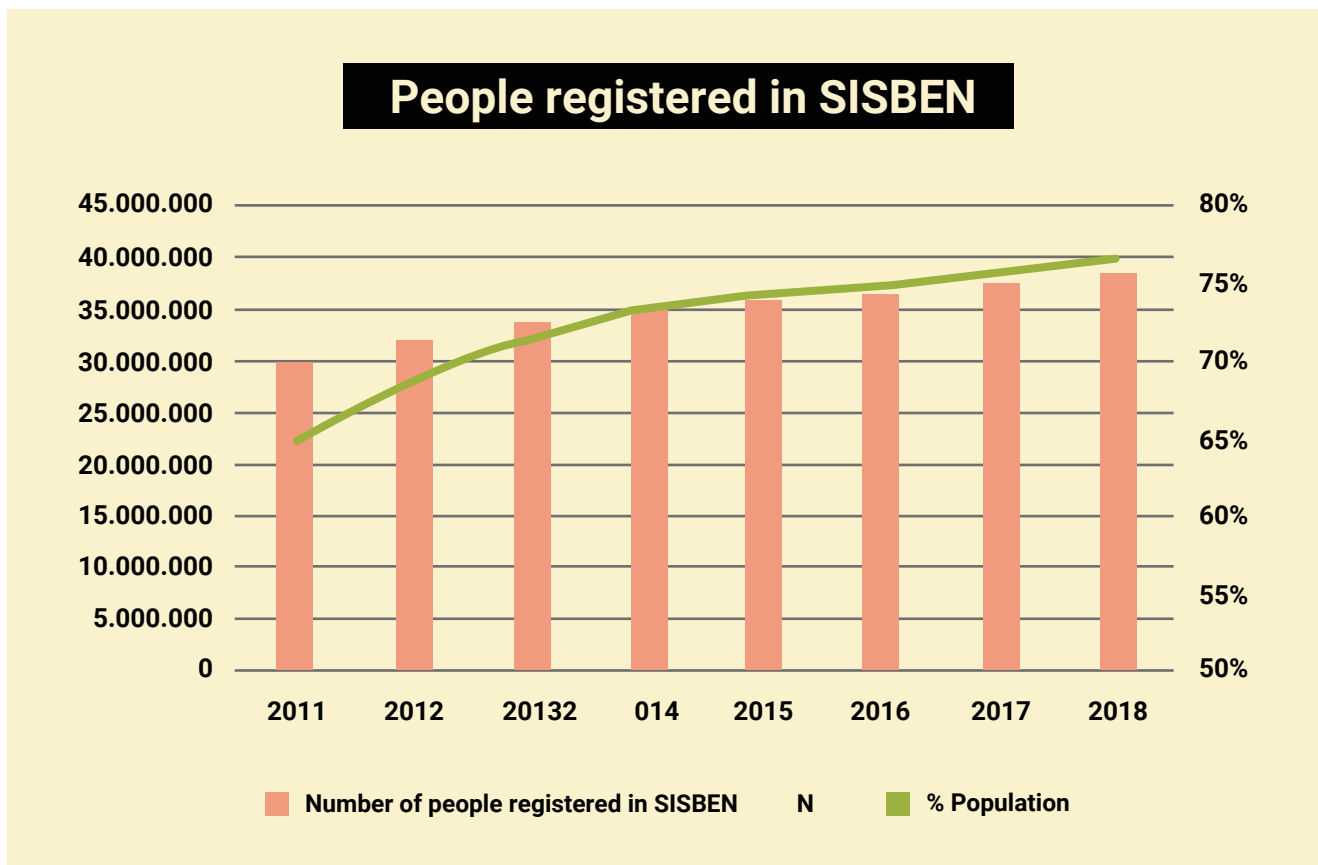
Table 1

0 10 20 30 40 50 60 70 80 90 100

The DNP announced it was targeting an expansion of its databases upon introducing the new SISBEN in 2017. Accordingly, it will collect more information about the population already in the system and, in turn, take data from other groups that were not in the areas where poverty had previously been located. There was talk, for this reason, of increasing the database to reach 40.5 million Colombians, that is, 80% of the population (Correa, 2017). Furthermore, with the incorporation of the income generation component, the survey questions increased considerably. It follows then that the SISBEN is moving now towards increasing the amount of information collected and the number of people in its records, even though they are not in the population segments that were generally in the database.

Figure 3:

Number of people registered in SISBEN as a percentage of the population.
 Source DNP, 2019.



The importance of the information received by SISBEN is illustrated by the emphasis the authorities placed on the negative consequences for those who provide false or inaccurate data. The new socioeconomic classification sheet used to survey potential beneficiaries includes two clauses that must be accepted to take the survey. These clauses specify that “the refusal to supply all the information requested will prevent your registration in SISBEN” and that the information is delivered under solemn oath, so “any alleged falsehood identified through database cross-checks will generate exclusion from SISBEN” and legal and judicial actions (National Planning Department, 2016).

These warnings show how DNP becomes more and more of an administrator of massive personal data through collecting consents based on the threat of denial of essential services. Furthermore, the targeting policy of social programs is evidently focused on looking for “gate-crashers” and threatening people rather than “looking for” them.

In 2017, the DNP issued Decree 441 that modified the guidelines to update and look for inconsistencies in the SISBEN databases. With this regulation, the institution was left in charge of the database validation processes and quality controls. Public entities will thereby be able to “make the information available without agreements to update and apply the validation and quality control processes” and will be able to enter into information exchange agreements with companies (National Planning Department DNP, 2017).

The Decree established two types of results of the validation processes through interoperability. On the one hand, “the exclusion” from the database is carried out upon the registered person’s death, by court order, or because of duplication of records. On the other hand, the “under verification” labeling where the DNP informs the territorial entity of the inconsistency. The territorial entity is thereupon supposed to inform the person of the situation and decide on excluding the records by an administrative act or, instead, request a new interview for reclassification. Within the six months following the DNP notice, the records’ exclusion will be communicated to the entities that manage social programs to withdraw the benefits (National Planning Department, DNP, 2017, Article 2.2.8.3.5).

The DNP determined nine grounds for placing the records in the SISBEN database “under verification”: (1) changes of residence without requesting a new survey, (2) registration of death in other databases, (3) unjustified change of information determined by the DNP, (4) record of income higher than the values determined by the DNP, (5) changes in socioeconomic conditions not reported, (6) reports from territorial entities, (7) reports from the entities that administer benefits, (8) inaccuracies or inconsistencies in the information, or (9) any other inconsistency considered by the DNP (National Planning Department DNP, 2017).

The SISBEN was presented from its implementation as the most “technical, objective, equitable and transparent instrument” to determine the poor “deserving” of receiving social benefits (McGee, 1999; Vélez et al., 1999). That said, the SISBEN, like any other measurement, embodies a conception of poverty, and, considering that the DNP is part of the government, it moves according to the needs of the current leadership. Indeed, for the State, it implies a particular political or social intent to point out who is in a situation of poverty (Menjura Murcia, 2016).



3. NEW TECHNOLOGIES AND EXPERIMENTATION WITH SOCIAL POLICY DATA.

As mentioned, the fourth version of SISBEN presented a series of changes where data analytics technologies are expected to help the State to be more “efficient” with public spending, looking for the “true” poor and punishing the “gate-crashers” within the system.

Next, we will analyze two cases that show the changes in logic and the experimentation processes with the SISBEN data.

a. The Experian case and financial profiling

One of the most interesting cases of information cross-checking is the use of databases from the financial risk rating agency Datacredito that belongs to the Irish multinational Experian.

In August 2018, DNP and Experian signed an Information Sharing Agreement to enable exchanging information between the two institutions and work for their common interest. In the case of DNP, the goal is to use the Experian financial risk database to “perform cross-checks with the information from the SISBEN database” and improve information quality by looking for inconsistencies and “thereby achieve greater precision regarding the solvency of the people surveyed (...) for the assignment of the scores” (National Planning Department & Experian Colombia, 2018).

The agreement establishes that Experian will “allow and facilitate” DNP access to two types of systems. On the one hand, access to information on the Quanto product, “a solution that allows estimating the income level of a natural person.” According to Experian (2019), a query in Quanto “yields a point estimate and the most likely income range of the person evaluated” at three levels (estimated value, lower and upper limit) according to the “risk to be accepted.” On the other hand, Experian undertook to provide DNP with access to the titleholders’ financial information in its databases. Likewise, the agreement mentions the usefulness of the “experience and knowledge in the development of analytical tools” for the operation of the DNP (National Planning Department & Experian Colombia, 2018).



The agreement also benefits Experian, as it was offered access to “non-confidential information” for the development of applications and services that banking institutions would use to facilitate access to credit for low-income people. In other words, the SISBEN information exchange is useful for Experian in building a credit risk profile of the most vulnerable people and developing more applications that aim to exploit their personal data. In this sense, “the DNP acts as a User and Source of information” for Experian (DNP, 2018 p. 5). Therefore, it is clear that the DNP’s priority is to establish surveillance models that ensure its databases’ quality rather than the consequences of these types of profiling models.

Likewise, Experian will not be liable for the quality and use of the information it delivers to the DNP. Always clarifying that Experian uses third-party data that make it technically impossible to guarantee their accuracy and integrity, since “Experian’s services involve services, models and techniques based on statistical analysis, probability and predictive behavior” (DNP & Experian, 2018, p. 10). The data that Experian shall deliver includes identification data such as approximate age range and gender; credit data such as the number of loans, loan amounts, and percentage of credit card usage; data from contracts with telecommunications companies such as the number of cellphone lines and the value of fees for device purchases (See Table 2).

Type of personal data	Data
Identification client	Geographic information
	Names and surnames
	Approx. age range
	Gender
	Number of debts
Loans	Type of loan
	Initial value of each loan
	Balance value of each loan
	Installment value of each loan
	Arrears of each loan
Credit cards	Number of credit cards
	Limit of each card
	Used value of each card
	Percentage of usage of each card
	Arrears on each credit card
Telecommunications sector	Number of cellphone lines
	Fee value of each line
	Arrears in fees for each line

Table 2:
Data present in the EXPERIAN Database that the DNP would use. (National Planning Department, 2019d).



b. The INNPULSA Colombia case: Big Data and innovation with data from Sisben people

The second technologies inclusion project into SISBEN was a call made at the beginning of 2019 by Innpulsa Colombia, a Trust with public resources of a private nature, with the goal of “selecting a company to create and execute a data analytics model that contributes to efforts to measure and detect fraud in Sisben.” The selected bidder must create and execute a data analytics model to apply Big Data with the support of the Massachusetts Institute of Technology (MIT), under consulting contract 015-2017, which allows answering the following questions:

- 1. What behavior patterns of the people registered in the SISBEN reflect characteristics that define them as prone to incurring fraudulent actions to alter the score to access benefits?*
- 2. What incentives drive alterations in the behavior of the people registered in the SISBEN? The above, to identify those users who intend to maintain their status as beneficiaries by altering the System’s characteristics. This analysis must be driven from variables identified in the available data.*
- 3. What trend analysis, of a combination of social programs and characteristics, make it possible to establish which users who are beneficiaries of a pathway of care in Sisben may be susceptible to social mobility?*
- 4. What are the most effective combinations of programs to overcome the vulnerability condition and exceed the score threshold to receive help from social programs? (FIDUCOLDEX, INNPULSA Colombia, & MINCIT, 2019, p. 16)*

The call is part of Contract 015-2017 with MIT to perform a diagnosis of Big Data in the country, designing a Big Data strategy from the State and designing a roadmap for “strengthening the Big Data industry in Colombia.” Pilot projects will be implemented for the second phase, during which Colombian companies will be selected to work with MIT with the goal of “stimulating private sector participation and strengthening this industry in the country (FIDUCOLDEX et al., 2019, p. 6). The execution of this project targets “knowledge transfer to the selected company through the support of MIT (...) to allow leveraging available data through the application of methodologies and techniques that reflect state of the art data usage.” Likewise, the regime applicable to these contracts “is solely under private law” (FIDUCOLDEX et al., 2019, p. 7).

According to the Terms of Reference, Big Data is defined through data such as “flows and sets resulting from the fingerprints left by human beings when using cell phones- (call records), credit cards (transactions), transportation (subway or bus and EZ Pass records), social media and search engines, or when their actions are recorded by sensors, even if physical.” This definition is relevant, considering that the data sources that are to be offered to the winning company include: SISBEN, Basic education enrollment, National Higher Education System (MinEducation), and conditional transfer systems. However, mention was made that the winning company and INNPULSA “may suggest and provide data from other sources that complement the analysis as long as it is possible to have access to the latter” (FIDUCOLDEX et al., 2019, p. 18).



The data analytics product to be delivered by the winning company of the tender has to construct an analysis of the data of the people registered in the SISBEN that includes “behavior patterns, the factors that alter it as a potential point of fraud” and in turn “the best combinations of social program packages in terms of efficiency to improve their situation as measured by the SISBEN score” (p. 17).

Data on people living in poverty as a resource for innovation

The call to use Big Data with SISBEN data shows a fundamental change regarding the management carried out by other governments. The Terms of Reference of the contract show a variation in the purpose of the SISBEN data that abandon the social objective and become active in promoting specific business models. In this sense, the contract coordinated by INNPULSA shows a trend in turning citizens’ personal data into assets for corporate innovation.

INNPULSA Colombia was created with the goal of promoting “entrepreneurship, innovation, and business strengthening” (FIDUCOLDEX et al., 2019, p. 2). For this reason, one of the focal points of its work connected with the objectives of the contract is:

Facilitate entrepreneurship and business innovation: designing and executing instruments, programs, specialized services, and financing strategies to accompany companies throughout their creation and growth process, from overcoming obstacles that prevent it, up to the management of resources and technical support so that they do not stop moving forward and conquering new markets as high impact companies FIDUCOLDEX et al., 2019, p. 5).

The INNPULSA contract is presented as part of a strategy to consolidate a data analytics market in Colombia (National Economic Policy Council, 2019, p. 21) and “turn Colombia into the Silicon Valley of Latin America” (Ocampo, May 14, 2019). In this case, the Colombian company selected would provide a service to the State while receiving MIT’s expert training and access to a database massive enough with which to experiment.

The treatment of people’s data is an elementary component of social justice in information systems (Heeks & Renken, 2016). In this case, according to planning sector regulations, “the National Planning Department² shall spell out (...) the information cross-checks necessary for data cleaning and updating, the guidelines for implementation and operation, and the design of methodologies” (National Planning Department DNP, 2017). The Decree also mentions that “the administration of Sisben and the information collected by it will be subject to the fundamental rights to equality, privacy, protection of personal data and the principles of transparency, morality, efficiency, quality, and publicity of the information, as well as all other governing the administrative function” (National Planning Department DNP, 2017). Likewise, the Decree specifies that the use of the SISBEN database by entities must be “based on the objectives and impacts pursued.” From this, it is illogical that the citizens’ data collected for social policy purposes are used for business growth and consolidation of a data exploitation market.

2. Sole Regulatory Decree from the National Planning Administrative Sector (Decree 1082 of 2015) modified by Decree 441 of 2017

Likewise, the SISBEN's data treatment is unknown by the citizenry who are classified. Through information access requests, we wanted to know about the contracts that INNPULSA had with MIT and the studies prior to the contracts, and the answer was that "the applicable regime for the acts and contracts concluded is exclusively private." Therefore, the information they handle was classified as reserved because it is "related to individuals' private information" under the Law of Habeas Data (Business Growth Management Unit, FIDULCOLDEX, MINCIT, & IN- NPULSA Colombia, 2019).

4. CONCLUSIONS

- The text pointed to reading the problem of a system like SISBEN from the perspective of data justice and critical analysis of technology. The main instrument for targeting social benefits has been presented historically as a "transparent, technical and equitable" solution that has been criticized for the factors that are taken into account in the measurement (Castañeda & Fernández, 2005; Cortes Nieto, 2012), but not as a problem of data exploitation and profiling of the vulnerable population. In this case, we see SISBEN as an individual population profiling system, based on data from multiple sources, the result of which is a score that simplifies the complex reality of poverty and classifies people to determine who "deserves" the protection of the State. Thus, the ideas of automation, interoperability, and Big Data take specific forms related to objectives, political interests, and these measurements' construction.
- When these technologies' discourses and objectives are analyzed, we see a system designed to find inconsistencies between the people registered and reduce the number of people who can access benefits. This is backed by a narrative of spending efficiency and better targeting of social benefits. Thus, we see a social policy that more than "looking for the poor," tries to "look for liars." In this manner, a design appears that uses technology to discard people and not to include historically excluded groups.
- The idea of the Welfare State as a system that "looks for" vulnerable individuals to score them and determine whether or not they deserve access to social rights has strengthened the surveillance processes on the population in an individualized and increasingly invasive way, in an attempt to focus resources among those who are "really" poor (Alson, 2019). This vision of social protection transforms the citizen, subject of rights, into a record that must be analyzed to be "worthy" of the State's protection (Jasanoff, 2017).



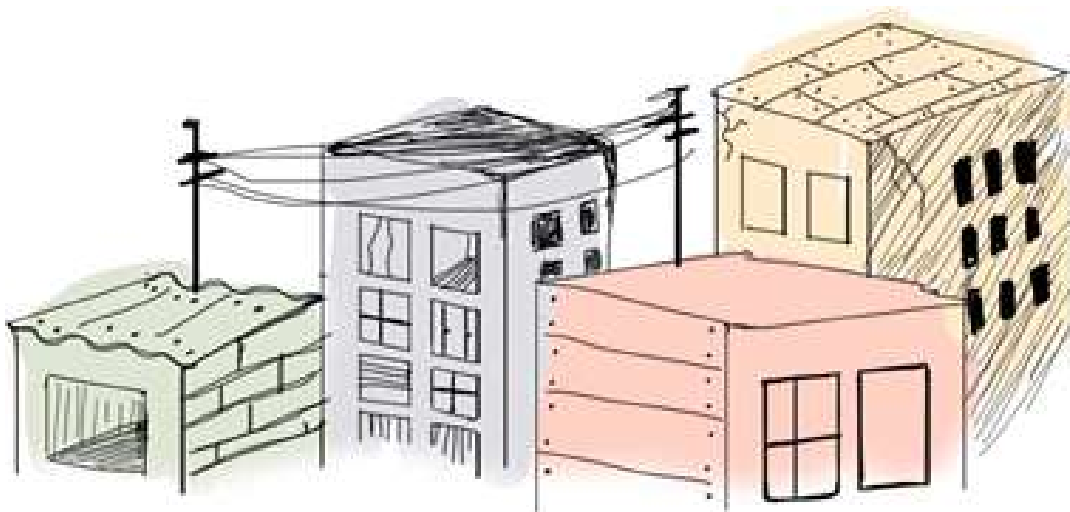
- Large information systems allow making and undoing the “deserving” through a system that, despite being arbitrary or unfair from its design (Constitutional Court, First Review Chamber, Sentence T-716/17, 2017), looks like the result of an objective examination of cutting-edge technologies.
- SISBEN narratives present issues of the State’s inability to reduce poverty in recent years as a situation of a technical rather than a political nature. Thus, the stagnation in poverty reduction in recent years and the great inequalities between regions is not the effect of failing social policies but of the lack of a more modern and precise instrument that can “search” for the “real” poor.
- The SISBEN IV rhetoric depicts “gate-crashers” as the central problem of poverty reduction languishing. Thus, surveillance and technology use come to the center as the most vulnerable link in a chain of construction of the SISBEN system that is susceptible to regional and national politicization (Castañeda & Fernández, 2005; Menjura Murcia, 2016).review
- As Arora (2019) mentioned, there is a tradition of the State to experiment with economically vulnerable people because the harms that can accrue are considered less important, and it is more difficult for them to access justice for reparations. Consequently, the SISBEN data become assets for consolidation of a digital market, and the beneficiaries become a space for experimentation regardless of the damage that may occur with these systems. The Innpulsa contract’s objectives do not recognize the public objectives of protecting people; instead, they seek to maximize resource efficiency and eliminate fraud. So this contract does not concern itself with the human costs of the system, and the damage of these experimentation processes on people’s social rights can be incalculable (Eubanks, 2018).
- The SISBEN case that we analyze in this text shows many of the limitations of regulations on personal data processing, transparency, and access to information by the State. People are obliged to give their consent to compare information with other databases, as they are threatened with losing their benefits. Thus, a policy that sows fear of losing benefits among people is exemplified, and they see the procedure for relating with the state as a risk.
- Likewise, people are graded by a system that uses their data for other purposes, which they do not understand and cannot challenge in case of impacts. The multiple times when information about peoples’ data treatment was denied, how they are graded, the studies, and the units of measure of the SISBEN algorithm show an opaque and unfair policy towards people.
- The changes and promises of the fourth version of SISBEN mark the narrative of the future of social assistance. A clear example is the Social Registry project, in which large-scale data analysis will be performed to determine the individual “merit” of all the social rights of people (Morales Manchego & Galindo Caballero, 2019). The project’s approach shows a State that, in an effort to determine the “truly” poor, tries to collect more data from people, placing vulnerable segments of the population in a state of surveillance and automating their relationship with the State to access their rights.
- For this reason, now more than ever, it is imperative to introduce concerns for privacy, dignity, and self-determination in the discussion of social protection to foster fairer systems that do not reproduce the historical inequities of our societies or make them acceptable as a result of futuristic technology.

5. RECOMMENDATIONS

- The design and use of the SISBEN must respond to the protection of people. The data collected, their treatment, and the consequences of the score can affect the citizenry's dignity and autonomy. In this case, we not only talk about a "measuring instrument" but also about a social classification system that affects people's lives.
- According to the Personal Data Protection Law and its principles of purpose and restricted circulation, the data collected in the SISBEN surveys are sensitive and should be used solely and exclusively for the purposes they were collected. This means that these data can only be used for the development of social policy and redistribution processes. The data may not be used for private profit or to promote business growth of data analytics in the country.
- People who are scored through SISBEN must have ways to be able to demand an explanation for the score they received. Likewise, in the case of exclusion, the reasons for any type of mark due to inconsistencies obtained must be explained, along with the databases used to detect it and the respective avenues to contest it available to people.
- It is necessary to perform data protection when the person in charge of the treatment is a State entity or public authority. Although personal data processing is part of many of the functions of entities or authorities, it is necessary to indicate clear criteria on the interpretation, specifically articles 5, 6, 10, and 13 of the Data Protection Law.
- Additionally, there is a need to strengthen the Personal Data Protection Division's investigative capabilities and sanction by the Attorney General's Office concerning the division of powers in article 23 of the same law.
- The SISBEN is an instrument that is key to the lives of millions of people, so the National Planning Department cannot be in charge of all the functions related to the system. In this case, the entity determines changes in the algorithm; it finances, monitors, and controls the system at the national level. The inclusion of other actors is essential to having a system balanced with the interests of a more equitable society that respects people's dignity.



- The idea of personal benefits requires data at the individual level, more oversight, and a pursuit focused on citizens who have been the target of the violence of marginalization. Thinking of alternatives to that individual basis might allow for different ways of measuring and qualifying poverty and not people, but rather geographic or social sectors that require state aid.
- The design of the oversight and control mechanisms of the SISBEN databases cannot focus on monitoring potential beneficiaries, building profiles, and predicting behaviors. These kinds of social policy predictions can mean excluding people from benefits and running counter to the principle of good faith, and the people whom SISBEN scores likewise include groups in a condition of vulnerability that cannot again be made targets of the violence of surveillance. The recommendation, on these grounds, is to consider surveillance and control mechanisms that focus on other links in the construction of SISBEN databases.
- The Social Registry's projects to replace the SISBEN ignore the risks of profiling and prediction for the exercise of social, civil, and political rights. For this reason, it is necessary to strike a balance between the dignity of people and the targeting processes of social records. The inclusion of technology should respond to the improvement of social benefits, not to the construction of a surveillance system on the beneficiaries. Besides, it is necessary to strike a balance between each social program's objectives and between the universal benefits and the targeted ones.
- The lack of participation of the people whom the SISBEN classifies excludes their experience and fails to recognize the difficulties that the system experiences in the field. The construction of the categories and the values included in the SISBEN algorithm must be a participatory process that involves people's experience and the expertise of other parts of the public sector and academia.



6. REFERENCES

Alson, P. (2019). Report of the Special rapporteur on extreme poverty and human rights: United Nations Human Rights Office of the High Commissioner. Retrieved from https://www.ohchr.org/Documents/Issues/Poverty/A_74_48037_AdvanceUneditedVersion.docx

Arbeláez Camaño, M. (2019). Sacar a los colados: objetivo de la encuesta para el nuevo Sisbén. RCN Radio. Retrieved from <https://www.rcnradio.com/colombia/sacar-los-colados-objective-of-the-survey-for-the-new-Sisben>

Arora, P. (2018). Politics of Algorithms, Indian Citizenship and the Colonial legacy, 22.

Arora, P. (2019). The next billion users: Digital life beyond the West: Harvard University Press.

Carnes, M., & Mares, I. (2015). Explaining the “return of the state” in middle-income countries: employment vulnerability, income, and preferences for social protection in Latin America. *Politics & Society*, 43 (4), 525–550.

Castañeda, T., & Fernández, L. (2005). Targeting social spending to the poor with proxy-means testing: Colombia’s SISBEN system. World bank human Development Network social protection unit discussion paper, 529.

Conexión Sur. (2019). Comunidad del área rural de Andes reacia a encuesta del Sisbén. Conexión Sur. Retrieved from <http://www.conexionsur.co/comunidad-del-area-rural-de-andes-reluctant-to-survey-of-Sisben/>

Consejo Nacional de Política Económica. (2019). CONPES 3975 Política Nacional para la Transformación Digital e Inteligencia Artificial. Bogotá.

Consejo Nacional de Política Económica, & Departamento Nacional de Planeación. (2016). DECLARACIÓN DE IMPORTANCIA ESTRATÉGICA DEL SISTEMA DE IDENTIFICACIÓN DE POTENCIALES BENEFICIARIOS (SISBÉN IV) (No. 3877).



Correa, MV (2017). 40,5 millones de colombianos estarán en 2019 en el Sisbén. El Colombiano. Retrieved from <https://www.elcolombiano.com/colombia/40-5-millones-de-Colombians-will-be-in-2019-in-el-Sisben-IA7741749>

Cortes Nieto, J. (2012). Pobreza, capacidades y medición. El intento de Colombia de medir las capacidades a través de una evaluación indirecta de los medios de vida: el SISBÉN. Revista Jurídica Piélagus, 11, 9–17.

DANE. (2019a). Empleo informal y Seguridad Social. DANE. Retrieved from <https://www.dane.gov.co/index.php/estadisticas-por-tema/mercado-laboral/empleo-informal-y-seguridad-social>

DANE. (2019b). Pobreza Monetaria y Multidimensional en Colombia 2018. DANE Información para todos.

Deacon, B. (2007). Global social policy and governance: Sage.

Departamento Nacional de Planeación. (2016). Ficha de caracterización socioeconómica -SISBÉN IV.

Departamento Nacional de Planeación. (2019a). Bases del Plan Nacional de Desarrollo (2018-2022): Pacto por Colombia, pacto por la equidad.

Departamento Nacional de Planeación. (2019b). Solicitud de Información No 20196000194942.

Departamento Nacional de Planeación. (2019c). Solicitud de Información No 20196000387092.

Departamento Nacional de Planeación. (2019d). Solicitud de Información No 20196000454702.

Departamento Nacional de Planeación, & Experian Colombia. (2018). Acuerdo de confidencialidad e intercambio de información entre el Departamento Nacional de Planeación y Experian Colombia.

Departamento Nacional de Planeación DNP. (2017). DECRETO 441 DEL 16 DE MARZO DE 2017

Departamento Nacional de Planeación DNP.

Departamento Nacional de Planeación-DNP. (2019). Respuesta solicitud de información Rad. 20196000194942.

Diario del Cauca. (2018). Piendamó pidió anular encuesta de Sisbén cuatro. Diario del Cauca. Retrieved from <https://diariodelcauca.com.co/noticias/local/piendamopidio-anular-encuesta-de-Sisben-four-447269>

Eubanks, V. (2018). Automating inequality: How high-tech tools profile, police, and punish the poor: St. Martin's Press.

Experian. (2019). QUANTO Datos de ingresos mensuales. Datacrédito Experian. Retrieved from <https://www.datacredito.com.co/empresas/index.jsp#/empresas/soluciones/quanto>

FIDUCOLDEX, INNPULSA Colombia, & MINCIT. (2019). TÉRMINOS DE REFERENCIA INNPULSA COLOMBIA "Términos de Referencia para seleccionar una empresa que cree y ejecute un modelo de analítica de datos que contribuya a los esfuerzos de medir y detectar fraudes en el Sisbén"BDPS-18.

Gupta, A. (2012). Red tape: Bureaucracy, structural violence, and poverty in India: Duke University Press.

Hall, A. (2007). Social Policies in the World Bank: Paradigms and Challenges. *Global Social Policy*, 7-151, 26. <https://doi.org/10.1177/1468018107078160>

Heeks, R., & Renken, J. (2016). Data justice for development: What would it mean? *Information Development*, 34 (1), 13. <https://doi.org/10.1177/0266666916678282>

Jasanoff, S. (2017). Virtual, visible, and actionable: Data assemblages and the sightlines of justice. *Big Data & Society*, 4 (2), 2053951717724477.

La Opinión. (2019). Sisbén: continuarán los que digan la verdad. La opinión. Retrieved from [# OP](https://www.laopinion.com.co/cucuta/Sisben-continuaran-los-que-digan-la-verdad-180221)
López Restrepo, A. (1995). Las etapas de la liberalización de la economía colombiana.

Lyon, D. (2009). Identifying citizens: ID cards as surveillance: *Polity*.

McGee, R. (1999). 'Technical, Objective, Equitable, and Uniform'? A Critique of the Colombian System for the Selection of Beneficiaries of Social Programs, *Sisben*.

Menjura Murcia, R. H. (2016). Los conceptos de pobreza en los instrumentos de focalización del Gasto Público en Colombia desde la última década siglo XX, alcances y perspectivas: El caso del Sisbén: Universidad Nacional de Colombia-Sede Bogotá.

Morales Manchego, M., & Galindo Caballero, M. (2019). Tras Sisbén 4 arranca era del registro social: director del DNP. Retrieved from <https://www.eltiempo.com/economia/sectores/tras-Sisbén-4-arranca-el-registro-social-para-reducir-y-focalizar-subsidios-421558>

Ocampo, Á. (14 de mayo de 2019). Duque: "Colombia se va a convertir en el Silicon Valley de América Latina". *La FM*. Retrieved from <https://www.lafm.com.co/tecnologia/duque-colombia-se-va-convertir-en-el-silicon-valley-de-america-latina>

Portafolio. (2017). Fueron 'sacados' más de 370.000 colados del Sisbén. *Portafolio*. Retrieved from <https://www.portafolio.co/economia/fueron-sacados-mas-de-370-000-colados-del-Sisbén-505550>

Sarmiento, A., González, J. I., & Rodríguez, L. A. (1999). Eficiencia horizontal y eficiencia vertical del Sistema de Selección de Beneficiarios (Sisbén).



Unidad de Gestión de Crecimiento Empresarial, FIDULCOLDEX, MINCIT, & iNNPULSA Colombia. (2019). Respuesta solicitud de información - "Ministerio de Industria, Comercio y Turismo" Radicado MINCIT No.:2-2019-010521- Radicado iNNpulsa Colombia No.:3949- Radicado FIDULCODELX S.A No.:64706409.

Vélez, C. E., Elkin Castaño, V., & Deutsch, R. (1999). Una interpretación económica del Sistema de Focalización de Programas Sociales: el caso Sisbén en Colombia.





EXPERIMENTING WITH POVERTY:

The SISBEN and data analytics
projects in Colombia

0 10 20 30 40 50 60 70 80 90 100

Fundación
Karisma
karisma.org.co

Twitter: @karisma
Facebook: @fundacionkarismaa
Instagram: @karismacol
YouTube: Fundación Karisma