The Journal of Sociology & Social Welfare

Volume 48
Issue 3 *Digital Social Work: Challenges, Trends and Best Practices*

Article 6

November 2021

Using Big Data to Manage Social Inclusion Programs

Esther Raya Diez
Universidad de La Rioja

Manuel Trujillo Carmona
Instituto de Estudios Sociales Avanzados

Domingo Carbonero Muñoz
Universidad de La Rioja

Using Big Data to Manage Social Inclusion Programs

Esther Raya Diez
Universidad de La Rioja
Facultad de Ciencias Jurídicas y Sociales

Manuel Trujillo Carmona
Instituto deEstudios Sociales Avanzados, IESA-CSIC

Domingo Carbonero Muñoz
Universidad de La Rioja

Technological developments based on Artificial Intelligence (AI) and empirical science in all areas of society are opening new opportunities for social work and social inclusion programs. AI relies on Big Data management systems, which in turn provide opportunities for descriptive inference and preventative measures, as well as data-informed decision making.

This article outlines the characteristics of Big Data and describes the process of designing a tool for diagnosing social exclusion, the SiSo scale. The tool consists of a scale that uses 25 variables to assess situations of social difficulty on the inclusion-exclusion spectrum. It is currently being used in the social services department of one of Spain's seventeen Autonomous Regions. The SiSo scale has the potential to advance the design of a Big Data system for social inclusion programs, provided we ensure the quality of the data. To this end, this study analyzes the suitability of the SiSo tool for measuring situations of social difficulty by conducting a Categorical Principal Components Analysis (CATPCA) and a Linear Principal Component Analysis.

The findings of the study confirm the tool's suitability and value for measuring levels of risk for social exclusion, as well as the feasibility of *implementing a system based on data generated by social inclusion programs. This article also highlights the opportunity that Big Data provides to generate knowledge by and for social work.*

Keywords: Artificial Intelligence, Digital Social Work, ICT, social exclusion, social services

Introduction

The information and knowledge society that emerged at the end of the twentieth century, characterized by the development of Information and Communication Technology (ICT) and globalization (Castells, 1996; Krüger, 2006), is giving way to a post-global society (Duguin, 2013, 2018; Benedetto, 2020) and new forms of relationships at the *local* level (Zuiker, 2010). In this new context, Artificial Intelligence (AI) is emerging as a technology with unprecedented capacity for social transformation.

AI refers to the simulation of human intelligence using machines and software (Pascual, 2019). Russel and Norvig (2009) identified four types of AI: systems that think like humans (artificial neural networks); systems that act like humans (robots); systems that use rational logic (expert systems); and systems that act rationally (intelligent agents). All these aspects are covered in the definition of AI used by the U.S. National Security Commission on Artificial Intelligence (NSCAI, 2021). According to the European Commission (2020), "AI is a collection of technologies that combine data, algorithms and computing power" (p. 2). AI is present in a large part of our daily activities (Lorente, 2017; Pombo et al., 2018) and has completely modified our lives (Zhongmei et al., 2020). Its various applications are used to address major challenges of advanced societies, such as treating chronic diseases, reducing traffic accidents, fighting climate change, and cybersecurity (European Commission, 2018). Progress using AI is also being made in social work and social services (López et al., 2020; Wilkerson et al., 2020).
Social work is a scientific discipline and professional practice aimed at social transformation that must be able to consciously and decisively integrate technological advances in research and all areas of professional practice. In this vein, e-social work gives rise to a new

specialized field that uses technology for social interventions (Castillo, 2017; Chan & Holosko, 2016; Coleman, 2011; López Peláez et al., 2018; Raya, 2018). The purpose of incorporating AI into social work practice is to strengthen the profession's capacity to protect citizens' rights; improve the quality of the services provided; increase professional, digital, and analog skills, and generate knowledge from practice. Nevertheless, social work practice seems oblivious to the potential of such technologies for social transformation (D'Antonio & de Lucas, 2017; Wilkerson et al., 2020), almost as if AI was something that pertains only to other disciplines (Raya, 2021). This fact was clearly evidenced by the management of social services during the COVID-19 pandemic (Wilkerson et al., 2020). Therefore, to advance in this respect would be to advance in the development of social work in the twenty-first century.

AI is developed through the management of large volumes of data. As highlighted by the Spanish Institute of Knowledge Engineering, meaning can be given to data for use in the decision-making process via AI (Instituto de Ingeniería del Conocimiento, 2021). Social work generates a large volume of data in the various fields in which professional practice takes place. By using AI systems, these data can be used to enhance social intervention processes.

In this article, we focus on the value of scientific knowledge generated by social workers and its application in social inclusion programs. This is accomplished through the creation of Big Data systems and the production and management of data stemming from social service interventions. This article describes the process of design and implementation of a tool to diagnose social exclusion resulting from collaboration between the Castilla-La Mancha Regional Government's Social Welfare Department and the University of La Rioja. Such collaboration was financed by the 2017-2020 European Social Fund for Castilla-La Mancha.

Castilla-La Mancha is one of Spain's seventeen Autonomous Regions. The region is located in the center of Spain, bordering to the south with Andalusia. According to the municipal census, Andalusia has 2,045,221 inhabitants, or 4.3% of the total population of Spain. The AROPE formula calculates the total number of persons below the poverty and social exclusion threshold. According to the formula, 30.7% of the population of Castilla-La Mancha is in a condition of poverty or social exclusion. This is higher than the national rate of 25.3%, and the rate of 21.4% for the 28 countries which belonged to the European Union at the time (EU-28).

Spain's decentralized social policy-making model gives a high level of autonomy to its regional governments. As a result, all Autonomous Regions have a Social Welfare Department that formulates social policies and interventions, including social inclusion programs. The creation and implementation of the SiSo *Situación Social* (Social Situation) scale is a response from Castilla-La Mancha's Social Welfare Department to demands of social service professionals and social inclusion organizations for valid and effective technological tools to assess complex situations. At the same time, the creation of this tool also responds to requests from various government agencies documenting social exclusion. Such reports are often needed, for example, to document eligibility for employment programs or discounts on energy bills.

The complex and ambiguous nature of social exclusion make it necessary for social work and social inclusion programs to have common conceptual frameworks, tools and language to help avoid situations of unfairness and injustice in the distribution of social services or benefits. Various authors have documented the need for valid and effective tools to assess social exclusion (Bramley & Bailey, 2018; Department of Equality, Justice, & Social Policies, 2013; Dermott & Main, 2018; Hernández, 2008; Gilbert, 2009; Gingrich & Lightman, 2015; Hernández; 2008; Laparra, 2008). The SiSo scale was created in response to that documented need. This name was chosen due to its neutral and non-stigmatizing nature.

Since its implementation in May 2018, the regular use of the SiSo tool by professionals in social inclusion programs has generated a large volume of data on all aspects related to social exclusion. These data, which are managed via Big Data applications, are extremely valuable for decision-making and social interventions. AI systems depend on and aim to produce high quality data. This fact motivated the literature review presented in this paper which served as a foundation for the development of the SiSo tool.

The main purpose of this article is to demonstrate the feasibility of developing Big Data systems based on data generated in social services, and describe the implementation process of the SiSo tool in one of Spain's seventeen Autonomous Regions. The first section of this article consists of a literature review on the role of Big Data for social inclusion. The second section describes the study's research methods. This includes a description of the SiSo tool's implementation process, the operational objectives and the methods used to check the quality of the tool. The study findings are presented in the third section, along with recommendations to improve the system based on statistical criteria. The conclusions section provides general guidelines as well as lessons learned from the implementation of the tool in Castilla-La Mancha.

Big Data for Social Inclusion

Big Data, or the storage and management of large datasets, is now a reality that touches many of our daily activities. Technological advancements have not only made storage easier, but have also enabled us to teach machines and enabled machines to learn by themselves. Through complex mathematical processes, data-based algorithms, new tools and software programs (Redondo, 2020), data are converted into information that can help in decision-making and generate greater competitiveness in organizations (Benavides Reina & Pedraza-Nájar, 2018; Contreras-Medina & Díaz-Nieto, 2014; Rodríguez et al., 2019).

The progress of Big Data in all areas of the economy, politics, and society at large is unstoppable (Duque-Jaramillo & Villa-Enciso, 2017; Duran, 2014; Mayer-Schönberger & Cukier, 2013; Van Rijmenam, 2014). Reportedly, there are three types of data: structured, unstructured, and semi-structured (Duran, 2014; Redondo, 2020). The first has a fixed format and uses numbers for storage in databases; in the second type, as the name suggests, data are disorganized and come from various sources. Finally, the third type represents a mix of the first two. The underlying idea of Big Data is that "the more you know about something, the greater your understanding, and the greater your ability to engage in informed decision-making aimed at finding solutions" (Redondo, 2020, p. 1). Advanced analytical tools can organize large amounts of data and can convert them into information needed for decision-making.

Big Data increases the possibility of generating knowledge for and through social work (Castillo, 2017; Coulton et al., 2015; Getz, 2014). To this end, data associated with social interventions need to be adequately managed, analyzed, and integrated. In the social services sector, data and information about users are collected at different times and through different systems. These data are often used to formulate public policies (Real & de las Heras, 2011; van Veenstra et al., 2020). The resulting databases and Information Management Systems facilitate access to relevant social intervention data. These systems, in turn, ensure that data are not dispersed across various applications, on paper or, worse, lost on shelves, in drawers, or in disorganized computer files.

The management system for data stored in databases or repositories must facilitate returning pertinent information to the individuals who supplied the data. After users have access to the data, they need to analyze them using selected criteria. In social inclusion programs, for instance, it is useful to create service user profiles, including their characteristics, and how they are similar or different. It is also useful to identify the variables associated with exclusion, and what population subgroups are vulnerable or in high-risk situations.

It has been suggested that in order to understand Big Data, we need to evaluate it using Big Data's Big Vs (Duran, 2014; Redondo, 2020; Van Rijmenam, 2014). These are volume, velocity, variety, veracity, value, and variability. These criteria must also be applied to data generated by social inclusion programs. The first V—volume, means that Big Data systems need large datasets. During social interventions, a wide range of data on service users and their families are collected at different points in time. These collected data make possible the creation of large and diverse datasets. For instance, since the SiSo tool was implemented in 2018, over 60,000 social service user data records have been created in Castilla-La Mancha.

The second V, velocity, refers to the speed at which data are received and processed. This requires the use of advanced information management systems. Such systems are dependent on those who supply the system with data. Many Big Data applications collect user data that results from internet browsing. In the case of social inclusion programs, however, Big Data depends on social work professionals to supply data to the system. Because of this, the effectiveness of Social Work Big Data systems depends on the intrinsic and extrinsic motivation of the professionals who provide the data independently from the cost or high level of investment made for the creation of information management system.

The third V refers to the variety of sources of data such as text, audio, and video that may require more intense processing. Social inclusion programs collect and process data on various areas of the service users' lives such as: economics, employment, housing, cohabitation, education, health, and other sociodemographics. Furthermore, the information management systems also keep track of data associated with the system's use, such as dates, time spent, etc. The system keeps track of user benefit requests and stores supporting documents related to such requests. This adds to the wealth of the Big Data system.

The fourth V, veracity, relates to questions of validity and reliability. We must avoid storing poor-quality data, given that this could lead to incorrect conclusions, decisions, or bias and discrimination (Buolamwini & Gebru, 2018; Degli-Esposti, 2019). Data must be valid and reliable in order to produce relevant outcomes. Veracity leads to the fifth V, value. Value is the ultimate goal of Big Data. Social inclusion programs seek the social integration or reincorporation of users that are vulnerable or at risk of exclusion. For this reason, the data supplied to the system must be useful towards this end. Additionally, it is important to differentiate between data relevant for direct client interventions and data that are useful for diagnosis, follow-up, and evaluation. Regardless of the type of data received, we must ensure that the resulting information is analyzed and returned to the professionals who supplied the data in a timely manner to make sure they do not lose their interest and motivation to provide data to the system.

The sixth V refers to variability. Large databases can be used for different purposes (Redondo, 2020). In the case of the SiSo tool, the data supplied to the system makes it possible to diagnose difficult social situations at the individual case level. In addition, the tool's dashboard provides updated information on the characteristics of the registered population based on different inclusion and/or exclusion variables. Furthermore, the SiSo tool can generate lists using up to five different filters. As a result, the system serves as a management tool for social intervention.

In addition to the above six Vs, we should also consider vulnerability, volatility, visualization, and validity (Van Rijmenam, 2014). Vulnerability refers to security and privacy issues; volatility refers to the data's obsolescence; visualization refers to the data or information's graphical representation, and validity refers to the value of the data.

The potential of Big Data is huge. Its main applications have been applied to business, banking and insurance (Management Solutions, 2015; Padilla-Barreto et al., 2017; Van Rijmenam, 2014, 2019). Big Data has been used as a prevention tool in medicine (Dash et al., 2019), to fight poverty (Pokhriyal & Jacques, 2017) and to assess child risk. The Allegheny Family Screening Tool (AFST), for instance, has been used to assess child risk. The AFST is a predictive risk modeling tool that rapidly integrates and analyzes hundreds of data elements for each individual involved in an allegation of child maltreatment (Vaithianathan et al., 2019). As in all fields in which Big Data is used, such as social work and social inclusion programs, careful data analysis and interpretation are needed in order for the system to represent a useful tool (Gillingham, 2020). The use of Big Data in social work has just begun, and judging by the low number of references linking social work to big data in journals indexed in Scopus, there is still a long way to go. A few authors that have written on this subject include He & Liu (2017), Gillingham & Graham (2017), and Fink (2018).

Methods

This section presents SiSo as a tool to collect data for its subsequent analysis, identifies the operational research objectives, and describes the procedures that were followed and the analyses that were performed.

Tool

The SiSo scale for measuring social inclusion is organized around six dimensions: economic, employment, education, housing, health, and relational. It is presented as a rubric with four levels of intensity with respect to social difficulty: little or no difficulty; some difficulty; considerable difficulty, and a lot of difficulty. Furthermore, the conceptual framework of the scale provides a description of the situations compatible with each level of difficulty. After the corresponding intervention interviews are conducted, professionals use the SiSo tool to record the levels of difficulty that best describe the various situations. This process usually takes between 7 and 10 minutes. SiSo then generates a report with a graphical presentation of the level of difficulty for the case in question. This provides a snapshot of the household situation at the time of the intervention. Subsequently, as many pertinent case reports as needed can be generated.

The data generated and collected are stored in the SiSo database. These data are subsequently analyzed to produce information useful for the formulation and evaluation of social policies and for the management of social inclusion programs. Since implementation of the tool in 2018 until January 2021, a total of 20,156 active files and 64,816 archived files with historical data have been stored in the database. These numbers should help us understand the potential for data generation of the social service delivery system.

After using the tool for three years, it has become necessary to assess the value of the collected data and evaluate its measurement capacity based on a mathematical model and an analysis of its principal components. This should enable us to identify possible design errors related to the selection of indicators and the development of the scale. The following operational research objectives seek to help us achieve this objective: (1) Compare the levels of difficulty resulting from the use of the scale through theoretical weighting and classify them using multivariate statistical tests, and (2) Propose a measurement system based on statistical criteria to strengthen the SiSo tool.

Procedure and Analyses

The performed analyses were divided into three stages. The first stage sought to create an initial indicator of social difficulty to describe the severity of the household's situation. Four levels of difficulty were created to describe the needs of service users, which we will describe in a subsequent section of this paper. The second stage sought to produce a second indicator by using different statistical analyses. This second indicator, in turn, has four different sections and it is used as a control in the next stage. The third and final stage sought to compare the two types of indicators created. The comparison of both types of indicators makes it possible to reconsider the values assigned to some of the variables, and the suitability of the variables used in the analysis.

First Classification: Building a First Level of Difficulty Indicator

The analyses of the first stage are aimed at creating an indicator of social difficulty based on weights given to selected variables. During the design of the scale, cut-off points in the definition, social exclusion, were determined based on a selected theoretical criterion. This was done by weighting the levels of difficulty for each corresponding dimension. It was therefore considered that structural factors related to exclusion such as economics, employment, and housing, should have a greater weight (0, 2, 4, and 6 points) than those relating to personal and education aspects (0, 1, 2, and 3 points). Factors related to health were given a midrange value (0, 2, 3, and 4 points). These criteria, used in other research studies conducted in Spain to measure social exclusion (Hernández, 2008; Sartu Federation, 2002), were used to establish levels of social difficulty.

While assessing different situations, a score is calculated by adding the scores corresponding to each of the 35 variables. This score ranges from 0 to 113. Furthermore, in order to obtain a more precise assessment, the overall level of difficulty is divided into four levels.

The client population was classified according to four levels of difficulty. The group with the lowest level of difficulty represents 3.2% (n=606) of the client population and includes users with scores of 28 points or lower. Two groups were assigned to the intermediate levels of difficulty. They represented 56.1% (n=10,644) and 38.4% (n=7,290) respectively of the total client population. Members of these two groups scored between 29 and 57 points, and between 58 and 85 points. Finally, the group with the highest level of difficulty represents 2.3% (n=428) of the client population. Its members scored 86 points or more (JCCM, 2018).

Second classification: The construction of a second indicator of level of difficulty using multivariate analysis techniques

In this stage, a Categorical Principal Components Analysis (CAT-PCA) and a Linear Principal Component Analysis (PCA) were used sequentially. The main reasons for using these analyses included the availability of nominal variables (Carbonero Muñoz & Ruíz Vega, 2016; Pérez et al., 2002). An attempt was made to understand if the variables adapted to the dimensions of the study, and to apply the principle of parsimony to reduce the number of study variables. As a result, six factorial scores, representing each of the six areas of the tool, were obtained. The PCA based on the factorial scores produced a metric variable that incorporated the various variables.

Lastly, four levels of difficulty were created. The lowest level of difficulty grouping 13.9% (n=2,628) of the client population included users with SD between -4.45 and -1. The intermediate levels of difficulty grouping 33.8% (n=6,406) and 39.8% (n=7,546) of the client population included users with SD between -1 and 0, and 0 and +1, respectively. Finally, the highest level of difficulty grouping 12.6% (n=2,393) of the client population included users with SD between +1 and 4.3.

Third stage: Comparison between the first and second classifications

At this stage of the study, the scores associated with the first and second indicators were compared. To this end, the following steps were taken: (1) Conducted linear correlation analysis between the objective first and second indicator scores; (2) Created graphical representation of the values obtained from the two indicators; and (3) Compared the two sets of indicator scores by performing basic descriptive analyses on the two scales.

Findings

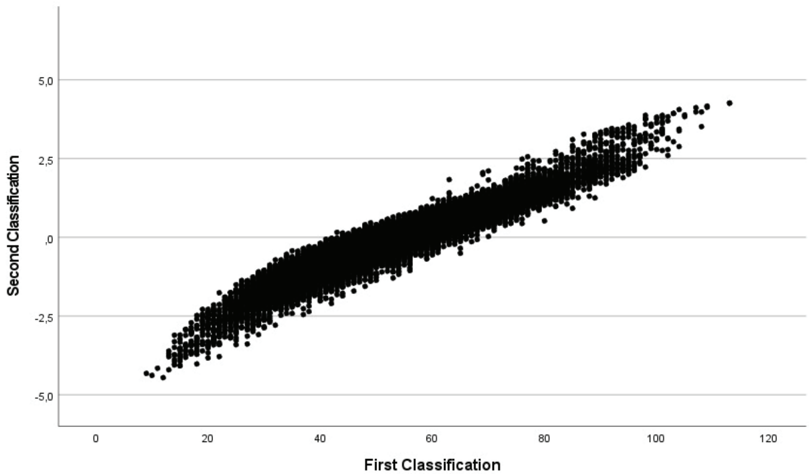
Linear correlation analysis and scatter diagram

The correlation between the two indicators aimed to identify the degree of significance and the strength of the association between both variables. This was done through a Pearson's correlation

(Pearson's R =.947; p<.001). The identified correlation coefficient and level of significance indicate a strong and significant correlation.

Furthermore, the findings of the statistical analysis were graphically illustrated by the following scatter diagram. The values resulting from the first and second classifications are represented on the X and Y axes, respectively. The results illustrated in the graph show that most cases clustered around the diagonal line.

Figure 1. Scatter Diagram Showing the Association between the First and Second Indicators of Social Difficulty

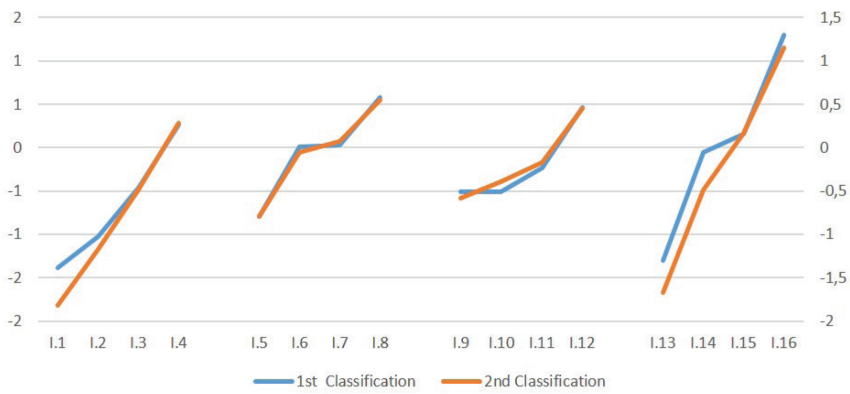


Comparison of Items According to Level of Difficulty Based on Two Indicators

The figures in the appendices show strong correlations for the scores for each dimension of social difficulty. The first groupings of social difficulty stemmed from the theoretical weighting and the second one stemmed from a statistical approach. Due to space limitations, we only report the most relevant findings related to the four dimensions: (1) economic; (2) employment; (3) housing; and (4) residential situation.

Findings related to the economic dimension show that users with incomes greater than 100% of Spain's median income and households that do not have material deprivation have a negative value in the second-stage indicator that is lower than their value in the first-stage indicator (Figure 1). This is reflected in variables 1 (Income level) and 4 (Home deprivation) in Figure 1. As mentioned above, items i.1 and i.13 have lower scores in relation to the second indicator of social difficulty.

Figure 2. Statistical Measurement of Economic Dimension (Z Scores)

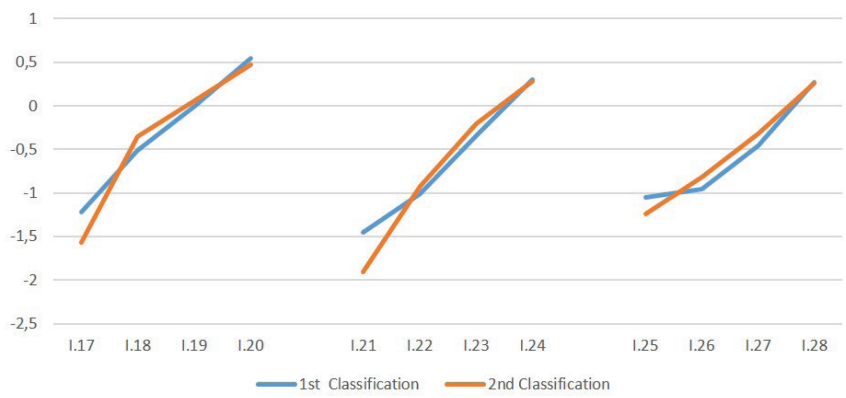


Note. Source: The authors

• V.1. Income level. I.1. Income greater than 100% of the country's median income; I.2. Income between 60% and 100%; I.3. Income between 30% and 60%; I.4. Income less than 40%;
• V.2. Sources of income. I.5. Income from non-contributory benefits. I,6. Income from the informal economy, non-recurring or family benefits; I.7. Income from the informal economy, non-periodic or family benefits. I.8. No income or marginal income.
• V.3. Income forecast. I.9. Income greater than one year; I.10. Income between 6 and 12 months; I.11. Income between 3 and 6 months; I.12. Without income or less than three months;
• V.4. Home deprivation. I.13. There is no deprivation; I.14, Less than 4 concepts; I.15. Absence of between 4-6 concepts; I.16. Absence of at least 7 concepts.

Something similar happens in relation to the employment and education dimensions. Users who do not have employment problems score somewhat lower on the second indicator of social difficulty. Similarly, the items related to up-to-date qualifications, actively searching for employment, and skills management, obtain lower values in the second statistical indicator than in the first weighted classification. Furthermore, the findings reported in Figure 3 show that the less severe items score a lower value in both dimension variables. This can be observed for variables 5, 6, and 7. For these variables, the values for items 1.17, i.21, and i.25 are lower in the classifications for the second indicator.

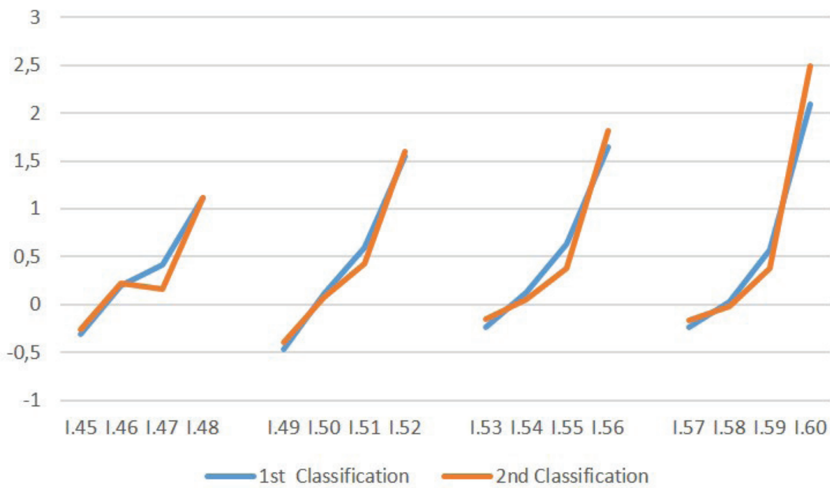
Figure 3. *Statistical Measurement of Employment Difficulties (Z Scores)*



Note. Source: The authors

• V.5. Employment situation. I.17. Marginal activities. Most of the adults of working age are long-term unemployed (+ 2 years). Inactive persons available for work who do not carry out any activity. I. 18. Unemployed in the last 2 years; Households where the main activity derives from an irregular/informal economy. I.19. Unstable employment and temporary occupations; Underemployment; and Inappropriate employment. I.20. No need to work and Activities for others/for themselves. No need to work
• V.6. Work intensity. I.21. Households in which those of working age did under three months of their total work potential in the last twelve months. I.22. More than half of those of working age did so between 3 and 6 months full time in the last twelve months. I.23. More than half of those of working age did so between 7 and 11 months full time in the last twelve months. I.24. More than half of those of working age worked full time during the last twelve months.
• V.7. Forecast of employment continuity with respect to the main job. I.25. More than 1 year. I.26. Between 7 and 12 months. I.27. Between 3 and 6 months. I.28. Unemployed or less than 3 months.

Figure 4. Statistical Measurement of Housing and Residential (Z Scores)



Note. Source: The authors

- V.12. Tenancy status. I. 45.Housing owned, assigned or rented with a guarantee for the cohabitation unit. I.46. Shared housing and/or sublease due to economic needs or Co-ownership with family conflict regarding the use of it. I.47. Accommodation with relatives for economic reasons and/or difficulties of access and staying in the home. I.48 Homeless, no home, eviction file and/or inadequate housing
- V.13. Housing conditions. I.49. Housing with adequate habitability conditions, basic equipment and supplies. I.50. There are some habitability deficiencies (it lacks fewer than three conditions); It lacks fewer than four pieces of equipment (<4) according to AROPE. I.51. Poor habitability conditions, Equipment deficiencies (between 4 and 8) according to AROPE. I.52. Lack of housing or roof, inadequate housing, Non-habitable housing (>6), Lack of pieces of equipment 4 (>8) according to AROPE.
- V.14. Access to housing. I.53. No barriers to the personal autonomy of household members. I.54. Barriers that do not affect mobility for the personal autonomy of household members. I.55. Barriers that limit mobility for the personal autonomy of household members. I56. Barriers that make mobility impossible for the personal autonomy of household members
- V.15. Location in the environment. I.57. Environment with a wide range of resources and public transport. I.58. Areas or neighborhoods with a low supply of resources and/ or communication. I..59. Disadvantaged, isolated and resource-poor environments. I.60. Illegal settlements, including lack of accommodation

In relation to housing, there is hardly any difference between the two indicators. It is worth mentioning that the items relating to illegal settlements, barriers that make mobility impossible, and a lack of equipment score higher in the new classification. This is especially noteworthy in the case of illegal settlements. Differences related to this can be observed on items i.52, i.56 e i.60.

A similar trend is observed in other analyzed dimensions, even though figures are not included due to space limitations. In the health dimension, low scores for "access to the healthcare system" variable in the first classification are noteworthy. The values obtained from the second indicator show that the "unsystematic use of the healthcare system" produced a lower score than "inappropriate use of the healthcare system." Likewise, the health dimension obtained a lower score in the PCA indicator given that it correlates less with the other dimensions. This fact causes the most extreme values to have a higher score in the first indicator than in the second indicator.

In the relational dimension, there are hardly any differences between the two indicators, although the more positive categories have somewhat lower values in the four variables. It should be noted that the family violence item has a lower score than the conflictive family relationships item. Findings primarily indicate that family violence is not associated with the most serious situations of social exclusion.

Comparison between the Classifications by Level of Difficulty

Comparisons were made in order to reassess the cut-off points used in the tool and to identify aspects needing correction. Despite a high degree of linear correlation, the findings below reveal similarities and differences according to the established cut-off points.

Table 1. Cross-Tabulation Classification

		First classification (Theoretical criterion)			
		Low level of difficulty	Low intermediate level of difficulty	High intermediate level of difficulty	High level of difficulty
Second classification (Statistical criterion)	Low level of difficulty	3.20%	10.70%	0%	0%
	Lower intermediate level of difficulty	0%	32.30%	1.40%	0%
	Higher intermediate level of difficulty	0%	13.10%	26.60%	0%
	High level of difficulty	0%	0%	10.40%	2.30%

Note. Source: The authors

In terms of similarities, analyses suggest that despite the different cut-off points, the extreme levels found in the first classification are also found in the second. Thus, 3.2% and 2.3% of low- and high-difficulty cases, respectively, are common to both. Significant differences, however, are found in relation to the two intermediate levels of difficulty. These differences are due to the different criteria associated with the respective cut-off points. Using the first indicator, 56.1% of cases were classified as having "low intermediate level of difficulty." Findings indicate that about a third of the cases (32.3%) are common to both indicator classifications. Furthermore, 10.7% of the users were classified as having "low level of difficulty" using the first indicator classification and having "low level of difficulty" using the second indicator classification. Similarly, 13.10% of cases were identified as having "low level of difficulty" using the first indicator classification and having "high intermediate level of difficulty" using the second indicator classification.

Using the first indicator, 38.4% are classified as having "high intermediate level of difficulty." However, 7 out of 10 cases with a "high intermediate level of difficulty" (70%) were common to both classifications. A total of 1.4% cases were classified as being in a situation of "high intermediate level of difficulty" according to the first indicator classification, while they were classified as having "low intermediate level of difficulty" in the second indicator classification. Finally, 3 out of 10 cases with "considerable difficulty," according to the first indicator classification (30%), were in the "high level of difficulty" category according to the second indicator classification (10.4%). Ultimately, the second indicator classification fully integrates the old extreme score groups and a significant percentage of the contiguous groups.

Comparison of Difficulty Levels by Mean, Standard Deviation, Variance, and Range

Findings suggest different options for creating cut-off points for both indicator classifications. The calculation was performed using standard deviations. This led to the creation of four client population groups. This calculation was intended to approximate the results using a common statistical criterion. Findings show similar distributions in the lower difficulty levels ("low level of difficulty," "low intermediate level of difficulty"), while also producing some differences in the higher difficulty levels ("high intermediate level of difficulty" and "high level of difficulty").

It should be noted that 34.2% of the client population in the first indicator classification and 39.7% of the client population in the second indicator classification were assigned to the "high intermediate level of difficulty." At the same time, 15.7% of the client population in the first indicator classification and 12.6% of the client population in the second indicator classification were assigned to the "high level of difficulty" category.

Moreover, findings of basic descriptive statistics indicate greater homogeneity and less dispersion in the values for the second indicator classification. The standard deviation attained the value of 14.19 in the first classification and a normalized value of 1 in the second classification.

Conclusions

This article has described the feasibility of producing scientific knowledge from social work practice via the development of Big Data information systems. These Big Data systems are necessary for Artificial Intelligence and both are currently being used in various fields to respond to social challenges. Big Data and AI can also strengthen the effectiveness of social work in defending citizen rights and improving the quality of services. This can be accomplished by using Big Data information systems with data generated by social work practitioners.

We have described the development process of the SiSo scale for measuring situations of social exclusion in the Autonomous region of Castilla-La Mancha in Spain. SiSo is a Big Data system that incorporates a large volume of systematized data. The quality of this tool was analyzed using CATPCA and PCA. The analysis showed a strong correlation between the scores of the theoretical construct indicator and the indicator resulting from a principal component analysis. This corroborates that the tool used represents a suitable means to measure social exclusion. Our analysis has also enabled us to identify needed improvements related to selected indicators.

The lessons learned as a result of the tool's implementation, and the verification of the quality of the data, should enable us to improve predictive analyses using Big Data. Predictive analyses will in turn enable us to identify variables most closely related to the different levels of exclusion and use them to perform systematic diagnoses of at-risk groups. Based on this experience, we recommend the addition of an S for system sustainability to the ten Vs of Big Data. In this study, sustainability was guaranteed by involving professionals in the different stages of the design, implementation, and monitoring. We also recommend the creation a single and centralized database, and providing professionals who supply data to the system with timely information stemming from the analysis of the data. This information could take the form of case or other reports they could use to engage in informed decision-making.

Social workers must use data science to support their interventions, and particularly to promote social inclusion. Furthermore, we should strengthen the relationship between social work, computer science, mathematics and other disciplines. We should use artificial intelligence and Big Data applications to open new horizons for the analysis of social problems and social interventions. We should remember that our ultimate goal is always to generate knowledge to enhance people's wellbeing.

Acknowledgments. We extend our most sincere thanks to all the professionals from Castilla-La Mancha in Spain and other Autonomous Regions who participated in the design and implementation of the SiSo scale and as a result contributed to the improvement of the tool.

References

Benavides Reina, M. R., & Pedraza-Nájar, X. L. (2018). *La gestión del conocimiento y su aporte a la competitividad en las organizaciones: Revisión sistemática de literatura* [Knowledge management and its contribution to competitiveness in organizations: A systematic literature review]. *Signos, 10*(2), 175–191. https://dialnet.unirioja.es/descarga/articulo/6726341.pdf

Benedetto, H. (2020). *Un mundo de contradicciones secundarias: ¿La post globalización?* [A world of secondary contradictions: Post globalization?] Mercosur Parliament. https://www.parlamentomercosur.org/innovaportal/v/12948/1/parlasur/un-mundo-de-contradicciones-secundarias:-la-post-globalizacion.html

Buolamwini, J., & Gebru, T. (2018). Gender Shades: Intersectional accuracy disparities in commercial gender classification, *Proceedings of the 1st Conference on Fairness, Accountability and Transparency, PMLR, 81*, 77–91. https://proceedings.mlr.press/v81/buolamwini18a/buolamwini18a.pdf

Bramley, G., & Bailey, N. (Eds.). (2018). *Poverty and social exclusion in the UK: The dimensions of disadvantage* (Vol. 2). Bristol University Press.

Carbonero Muñoz, D., & Ruíz Vega, A. (2016). *Evolución en los procesos e itinerarios de las personas sin hogar en España: ¿Retorno al pasado?* [Evolution in the processes and itineraries of homeless people in Spain: A return to the past?]. *Empiria, 34*, 39–78. https://doi.org/10.5944/empiria.34.2016.16522

Castells, M.L. (1996). *La era de la información: Economía, sociedad y cultura.* [The information age: Economy, society and culture] (Vol.1). Alianza Editorial.

Castillo, J. (2017) *El trabajo social ante el reto de la transformación digital. Big data y redes sociales para la investigación y la intervención social* [Social work facing the challenge of digital transformation. Big data and social networks for research and social intervention]. Thomson Reuters Aranzadi.

Chan, C., & Holosko, M. J. (2016). A review of information and communication technology enhanced social work interventions. *Research on Social Work Practice*, *26*(1), 88–100. http://dx.doi.org/10.1177/1049731515578884

Coleman, N. (2011). *E-social work: A preliminary examination of social services contact centres* [Unpublished doctoral dissertation]. University of Warwick. http://wrap.warwick.ac.uk/51364/ Contreras-Medina, D. I., & Díaz-Nieto, E. S. (2014). *La gestión del conocimiento factor clave de competitividad. Un estudio de los modelos y paradigmas* [Knowledge management as a key factor of competitiveness. A study of models and paradigms]. *Contributions to Economics*. https://www.eumed.net/ce/2014/2/conocimiento-competitividad.html

Coulton, C. J., Goerge, R., Putnam-Hornstein, E., & de Haan, B. (2015). *Harnessing big data for social good: A grand challenge for social work*. American Academy of Social Work & Social Welfare. https://grandchallengesforsocialwork.org/wp-content/uploads/2015/12/WP11-with-cover.pdfD'Antonio, S., & de Lucas, F. (2017). *Trabajo social y tecnología: Acomodación distante y precariedad* [Social work and technology: Distant accommodation and precariousness]. In E. Raya (Ed.), *Innovación social en la práctica del trabajo social.*Tirant Humanidades.

Dash, S., Shakyawar, S. K., Sharma, M., & Kaushik, S. (2019). Big data in healthcare: Management, analysis and future prospects. *J Big Data, 6*(54), 1–25. https://doi.org/10.1186/s40537-019-0217-0

Degli-Esposti, S. (2019). *Los algoritmos también discriminan a los seres humanos: Nosotros podemos impedirlo* [Algorithms also discriminate against humans: We can prevent it]. *The Conversation.* https://theconversation.com/los-algoritmos-tambien-discriminan-a-los-seres-humanos-nosotros-podemos-impedirlo-124794

Department of Equality, Justice, and Social Policies. (2013). *Instrumentos comunes de diagnóstico social y valoración de la exclusión* [Common tools for social diagnosis and assessment of exclusion]. Basque Government. https://www.euskadi.eus/instrumentos-comunes-diagnostico-social-valoracion-exclusion/web01-a2gizar/es/

Dermott, E., & Main, G. (Eds.). (2018). *Poverty and social exclusion in the UK: The nature and extent of the problem* (Vol. 1). Bristol University Press.

Duguin, A. (2013). *La cuarta teoría política* [The fourth political theory]. Ediciones Fides,.

Duguin, A. (2018). *El auge de la cuarta teoría política: La cuarta teoría política* [The rise of the fourth political theory: The fourth political theory] (Vol. 2). Ediciones Fides.

Duque-Jaramillo, J., & Villa-Enciso, E. (2017). *Big Data: Desarrollo, avance y aplicación en las organizaciones de la era de la información* [Big data: Development, advancement and application in information age organizations]. *Revista CEA, 2*(4), 27–45. http://dx.doi.org/10.22430/24223182.169

Duran, X. (2014). *El imperio de los datos. El Big Data, la privacidad y la sociedad del future* [The empire of data. Big Data, privacy and the society of the future]. Universidad de Valencia.

European Commission (2018). Coordinated Plan on Artificial Intelligence, Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Codmmittee and the Committee of the Regions. https://eur-lex.europa.eu/legal-content/ES/TXT/HTML/?uri=CELEX:52018DC0795&from=DA

European Commission (2020). White paper on artificial intelligence, CO (2020) 65 final, https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf

Fink, A. (2018). Bigger data, less wisdom: The need for more inclusive collective intelligence in social service provision. *AI & Society*, *33*(1), 61-70. https://doi.org/10.1007/s00146-017-0719-2

Getz, L. (2014). Big data's impact on social services. *Social Work Today*, *14*(2), 28. https://www.socialworktoday.com/archive/031714p28.shtml

Gilbert, N. (2009). European measures of poverty and "social exclusion": Material deprivation, consumption, and life satisfaction. *Journal of Policy Analysis and Management*, *28*(4), 738-744. https://doi.org/10.1002/pam.20471

Gillingham, P. (2020). The development of algorithmically based decision-making systems in children's protective services: Is administrative data good enough? *The British Journal of Social Work*, *50*(2), 565-580. https://doi.org/10.1093/bjsw/bcz157

Gillingham, P., & Graham, T. (2017). Big data in social welfare: The development of a critical perspective on social work's latest "electronic turn". *Australian Social Work, 70*(2), 135–147. https://doi.org/10.1080/0312407X.2015.1134606

Gingrich, L.G., & Lightman, N. (2015). The empirical measurement of a theoretical concept: Tracing social exclusion among racial minority and migrant groups in Canada. *Social Inclusion, 3*(4), 98–111. https://doi.org/10.17645/si.v3i4.144

He, X., & Liu, P. (2017, July 21-24). *Empirical study on the construction of a unified information system with big data analysis: A case of social work organizations and civil affairs departments* [Paper presentation]. 2017 IEEE International Conference on Computational Science and Engineering; IEEE/IFIP International Conference on Embedded and Ubiquitous Computing, Guangzhou, China. https://doi.org/10.1109/CSE-EUC.2017.48

Hernández, M. (2008). *Exclusión social y desigualdad* [Social exclusion and inequality]. University of Murcia.Instituto de Ingeniería del Conocimiento. (2021). Big data y inteligencia artificial [Big data and artificial intelligence]. https://www.iic.uam.es/big-data-inteligencia-artificial/

JCCM (2018) Manual de procedimiento de la herramienta SiSo. [SiSo Tool Procedure Manual]. Universidad de La Rioja, Consejería de Bienestar Social de Castilla-La Mancha.

Krüger, K. (2006). El concepto de 'sociedad del conocimiento' [The concept of the 'knowledge society']. *Revista Bibliográfica de Geografía y Ciencias Sociales,, 6*(683). http://www.ub.es/geocrit/b3w-683.htm

Laparra NM. (2008). La configuración del espacio social de la exclusión en España" [The configuration of the social space of exclusion in Spain]. In M. Laparra & B. Pérez (Eds.), *Exclusión social en España: Un espacio diverso y disperso en intensa transformación* (pp. 405–423). FOESSA Foundation.

López Peláez, A., Pérez Garcia, R., & Aguilar-Tablada Massó, Mª.V. (2018). E-social work: Building a new field of specialization in social work? *European Journal of Social Work, 21*(6), 804–823. http://dx.doi.org/10.1080/13691457.2017.1399256

López, B., Bermúdez, L.M., & Pascual, J. (2020). *Buena práctica: SisVAT-COVID19 sistema de vigilancia y alerta temprana en el sistema Asturiano de servicios sociales* [Best practice: SisVAT-COVID19 monitoring and early warning system in the Asturian social services system]. *socialasturias.es.* http://bit.ly/SisVAT-COVID19

Lorente, A. (2017, July 25). *Inteligencia artificial for good* [Artificial intelligence for good]. Think Big. https://blogthinkbig.com/inteligencia-artificial-for-good

Management Solutions (2015). *Data science y la transformación del sector financiero* [Data science and the transformation of the financial sector]. https://www.managementsolutions.com/sites/default/files/publicaciones/esp/Data-Science.pdf

Mayer-Schönberger, V., & Cukier, K. (2013). *Big data: La revolución de los datos masivos* [Big data: A revolution that will transform how we live, work, and think]. Turner.

National Security Commission on Artificial Intelligence. (2021). *FAQs.* https://www.nscai.gov/about/faq/

Padilla-Barreto, A., Guillén, M., & Bolancé, C. (2017). *Big-Data analytics in seguros. Revista Anales del Instituto de Actuarios Españoles* [Big-data analytics in insurance. Annals of the Institute of Spanish Actuaries], *Anales de Instituto de Actuarios Españoles, 4*(23), 1–19. https://www.actuarios.org/wp-content/uploads/2017/11/1_19_ANALES_1_Final.pdf

Pascual, J. A. (2019). *Inteligencia artificial: Qué es, cómo funciona y para qué se utiliza en la actualidad* [Artificial intelligence: What is it, how does it work and what is it being used for?]. Computer Hoy. https://computerhoy.com/reportajes/tecnologia/inteligencia-artificial-469917

Pérez, M., Sáez, H., & Trujillo, M. (2002). *Pobreza y exclusión social en Andalucía* [Poverty and social exclusion in Andalusia]. Editorial Politeya.

Pokhriyal, N., & Jacques, D. C. (2017). Combining disparate data sources for improved poverty prediction and mapping. *Proceedings of the National Academy of Sciences of the United Stated of America, 114*(46), E9783-E9792. https://doi.org/10.1073/pnas.1700319114

Pombo, C., Gupta, R., & Stankovic, M. (2018). *¿Quieres automatizar un servicio social con inteligencia artificial? 4 consideraciones clave* [Want to automate social services with artificial intelligence? 4 key considerations]. BID Mejorando Vidas. https://blogs.iadb.org/conocimiento-abierto/es/servicio-social-con-inteligencia-artificial/

Raya, E. (2006). *Indicadores de exclusión social: Una aproximación al estudio aplicado de la exclusión social* [Social exclusion indicators: An approach to the applied study of social exculsion]. Universidad del País Vasco.

Raya, E. (2018). E-inclusion and e-social work: New technologies at the service of social intervention. *European Journal of Social Work, 21*(6), 916–929. https://doi.org/10.1080/13691457.2018.1469472

Raya, E. (2021). Inteligencia artificial y srabajo Social: Aportaciones para una IA ética al servicio de las personas [Artificial intelligence and social work: Contributions for an ethical AI at the service of people]. In G. Kirwan & A. López Peláez (Eds..), *Handbook of digital social work*. Routledge.

Real, M. J., & de las Heras, V. (2011). MEDAS, una apuesta por las nuevas tecnologías y la calidad en la práctica del trabajo social en los servicios sociales [MEDAS, a commitment to new technologies and quality in the practice of social work in social services]. *Bits: Boletín Informativo trabajo social, 15.* https://dialnet.unirioja.es/servlet/articulo?codigo=3711899

Redondo, B. (2020). *Big data: ¿Qué es y cómo funciona?* [Big data: What is it and how does it work?]. Mailjet. https://es.mailjet.com/blog/news/big-data/

Rodríguez, V., Munuera, P., Raya, E., & Lascorz, A. (2019). *Instrumentos de valoración, diagnóstico y evaluación en trabajo social* [Assessment, diagnosis, and evaluation instruments in social work]. In E. Sobre monté & A. Rodríguez (Eds.), *El trabajo social en un mundo en transformación ¿Distintas realidades o nuevos relatos para la intervención?* Tirant Lo Blanch.

Russel, S., & Norvig, P. (2009). *Inteligencia artificial: Un enfoque moderno* [Artificial intelligence: A modern approach]. Pearson Prentice Hall.

Vaithianathan, R., Kulick, E., Putnam-Hornstein, E., & Benavides Prado, D. (2019). *Allegheny family screening tool: Methodology, version 2.* Allegheny County Department of Human Services. https://www.alleghenycountyanalytics.us/wp-content/uploads/2019/05/Methodology-V2-from-16-ACDHS-26_PredictiveRisk_Package_050119_FINAL-7.pdf

Van Rijmenam, M. (2014). *Think bigger: Developing a successful big data strategy for your business*. Amacom.

Van Rijmenam, M. (2019). *The organisation of tomorrow: How AI, blockchain and analytics turn your business into a data organization*. Routledge.

van Veenstra, A. F., Grommé, F., & Djafari, S. (2020). The use of public sector data analytics in the Netherlands. *Transforming Government: People, Process and Policy*. http://dx.doi.org/10.1108/TG-09-2019-0095

Wilkerson, D. A., Wolfe-Taylor, S. N., Deck, C. K., Wahler, E. A., & Davis, T. S. (2020). *Social Work Education*, *39*(8), 1137–1145. https://doi.org/10.1080/02615479.2020.1807926

Zhongmei, L., Yu-Che, H., & Cui, B. (2020, March 6-9). *A study for application research of 5G data acquisition and testing* [Paper presentation]. 2020 5th IEEE International Conference on Big Data Analytics, Xiamen, China. https://doi.org/10.1109/ICBDA49040.2020.9101298

Zuiker, S. (2010). Think global/Design glocal. *The magazine for managers of change in education*, *50*(5), 37–40.https://www.researchgate.net/publication/277870861_Think_Global_Design_Glocal