

# Dirección de Desarrollo Digital

Unidad de Científicos  
de Datos



**El futuro  
es de todos**

**DNP**  
Departamento  
Nacional de Planeación



## ARTICULACIÓN DE LOS PLANES DE DESARROLLO TERRITORIAL 2020-2023 CON LOS OBJETIVOS DE DESARROLLO SOSTENIBLE Y EL PLAN NACIONAL DE DESARROLLO 2018-2022

### Entidad

Departamento Nacional de Planeación

- Dirección de Desarrollo Digital.
- Dirección de Seguimiento y Evaluación de Políticas Públicas

### Sector

Territorial

### Lenguaje

Python y R

### Fuente de datos

Planes de Gobierno Territoriales y tabla de Indicadores ODS y PND tomada de Planes de Inversión. Páginas oficiales de Naciones Unidas. Plan Nacional de Desarrollo

### Presentación

Los Planes de Desarrollo Territorial (PDT) son los instrumentos de planificación con los cuales las administraciones locales definen los programas y proyectos que ejecutarán durante su período de gobierno. Dada la importancia estratégica de los PDT, es de interés nacional que su formulación sea acorde con las estrategias de país y con los lineamientos establecidos en el Plan Nacional de Desarrollo (PND). Al hacer uso de metodologías de minería de texto, este proyecto permite identificar los temas de mayor relevancia en los PDT, estimar el nivel de alineación de los PDT con los ODS y estimar el nivel de alineación de los PDT con lo establecido en el PND 2018-2022.

*Territorial development plans (PDTs, for its Spanish acronym) are the planning instruments with which local administrations define the programs and projects that they will execute during their government period. Given the strategic importance of PDTs, their alignment with the national strategy and with the guidelines established in the National Development Plan are of great importance. Therefore, this project identifies the most relevant subjects contained in the PDTs, estimates the level of alignment of the PDTs with the ODS, and the level of alignment of the PDTs with the National Development Plan 2018-2022, by using text mining methodologies.*

### Objetivo general

Analizar mediante un proceso automático la articulación de los Planes de Desarrollo Territorial (PDT) 2020 con los Objetivos de Desarrollo Sostenible (ODS) y el Plan Nacional de Desarrollo (PND) 2018 – 2022.

### Objetivos específicos

1. Establecer una medida de similitud estimada entre cada uno de los PDT 2020 y los ODS y entre cada uno de los PDT 2020 y los pactos del PND 2018-2022.
2. Identificar las temáticas más importantes de los programas de gobierno mediante un análisis descriptivo de los textos.
3. Facilitar la interpretación de los resultados mediante una herramienta de visualización.



## Metodología

El desarrollo de este proyecto se dividió en cinco fases; la lectura automática y limpieza de los textos, el análisis descriptivo, el cálculo de medidas de similitud, el análisis de indicadores ODS y PND, y la consolidación de los resultados. El detalle de cada etapa se muestra a continuación.

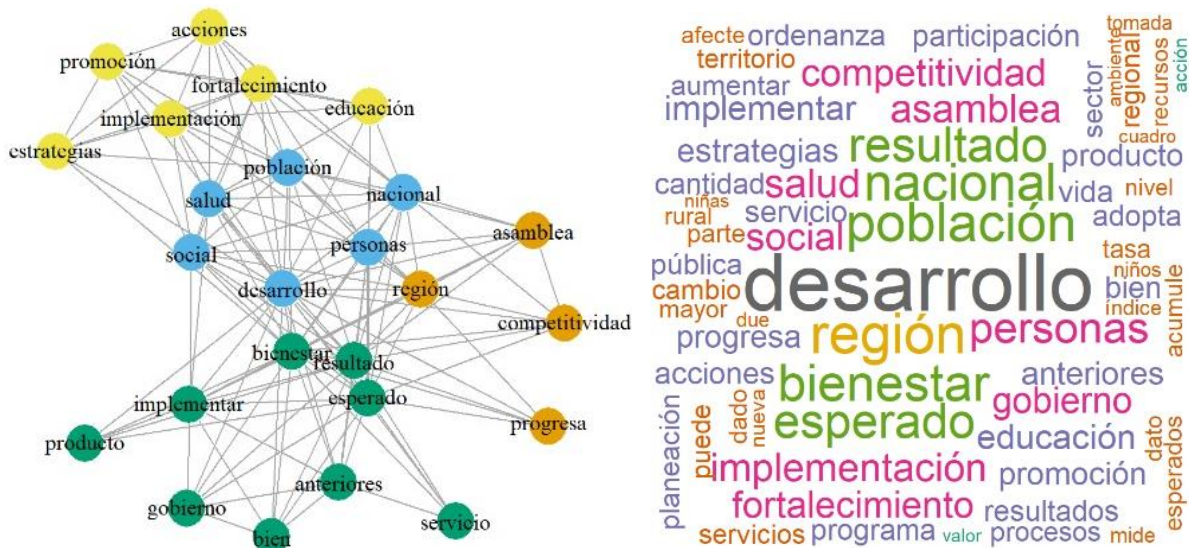
- **Lectura automática y limpieza de textos**

Para el análisis descriptivo de los planes de gobierno se realizó la lectura automática de los PDF utilizando Python, con lo que se obtuvo una cadena de texto del contenido de las páginas de cada plan. Algunos documentos requirieron del paso por un OCR (reconocimiento óptico de caracteres), un paso necesario pero que puede generar imprecisiones en la lectura de algunos caracteres. Una vez obtenidas las cadenas de texto de cada plan, se ejecutó la limpieza de las mismas con R. Dicha limpieza consistió en la transformación del texto a minúsculas y en la remoción de números, signos de puntuación y demás caracteres que no fueran letras; también se removieron conectores, preposiciones y palabras que no agregan significado al texto, entre las cuales se incluyeron nombres no ambiguos de las entidades territoriales y nombres no ambiguos de personas. Para este proyecto no se realizó la remoción de tildes, pues la inclusión de palabras como “género” (que sin tilde se relacionaría con “generar”) era relevante en este análisis. Se resalta, sin embargo, que los acentos son una de las principales limitaciones que se tienen al utilizar un OCR para leer los documentos. El resultado de este tratamiento es lo que se denominará “texto limpio”.

- **Análisis descriptivo**

Habiendo completado la lectura y limpieza de los textos, se realizó un conteo de palabras (unigramas) para identificar cuáles aparecían con mayor frecuencia, resultado que se utilizó como indicador de los temas principales que se tratan en cada plan de gobierno. Adicionalmente, se realizó un análisis de coocurrencia de términos, que es un conteo de las veces que aparece cada posible par de términos en una misma página. Con base en este análisis de coocurrencia se construyó una red de palabras agrupada, la cual ilustra que tan relacionadas se encuentran las palabras más mencionadas y agrupa dichos términos según el nivel de coocurrencia. Los grupos generados permiten identificar posibles temas centrales de cada plan de gobierno. La Figura 1 presenta un ejemplo de una red de palabras agrupada y una nube de palabras, elementos utilizados para presentar los resultados del análisis descriptivo.

Figura 1: Red de palabras agrupada (izquierda) y nube de palabras (derecha) construidas con base en un plan de gobierno.



Fuente: Elaboración Propia.



• **Cálculo de medidas de similitud**

Para la tercera fase del proyecto se comenzó consolidando, en cadenas únicas, los planes de gobierno, los textos disponibles de cada ODS y los pactos del PND. En el caso de los ODS, se juntaron diferentes tipos de texto, que fueron (1) los textos descriptivos y datos destacables de cada ODS<sup>1</sup>; (2) los documentos sobre “¿Por qué es importante?” relacionados con cada ODS<sup>2</sup>; (3) el detalle de objetivo, metas e indicadores, contenido en la base de datos que se ha utilizado como insumo en proyectos anteriores sobre ODS, y (4) las palabras clave que se identificaron en el proyecto “Rastreo de recursos para los ODS”<sup>3</sup>. Para este análisis se excluyó el ODS 17 (alianzas para alcanzar los objetivos) ya que este es transversal a los demás objetivos e involucra relaciones en el ámbito nacional e internacional, lo que dificulta su caracterización e identificación.

Habiendo obtenido las cadenas de texto consolidadas, se realizó una vectorización (es decir, una representación numérica) de los textos de cada ODS y cada pacto del PND. Para ello, se utilizó el modelo de bolsa de palabras (*bag of words*), que consiste en construir una matriz con palabras que pertenecen al corpus de los documentos y son únicas por columna. En dicha matriz, las filas muestran el número de veces que aparece cada palabra en un determinado objetivo o pacto, según corresponda. Por ejemplo, si un ODS fuera únicamente “acabar pobreza extrema”, las columnas “acabar”, “pobreza” y “extrema” tendrían un valor de 1 en la fila de la matriz, mientras las demás columnas tendrían un valor de 0. Otro ejemplo se presenta en la tabla 1, allí se tienen dos cadenas de texto; “Tasa de ocupación” y “Tasa de formalidad laboral”, las columnas están compuestas por las palabras contenidas en los textos y cada texto corresponde a un vector fila.

*Tabla 1: Representación numérica de textos con el modelo de bolsa de palabras.*

	de	ocupación	formalidad	tasa	laboral
Tasa de ocupación	1	1	0	1	0
Tasa de formalidad laboral	1	0	1	1	1

*Fuente: Elaboración propia.*

Habiendo realizado la vectorización de los textos ODS y PND, se procedió a vectorizar los planes de gobierno con base en el conjunto de palabras presentes en los textos ODS y PND. Esto es esencial debido a que los vectores finales necesitan estar en el mismo espacio vectorial para ser comparables. En términos prácticos, la realización de este procedimiento sería equivalente a, por un lado, eliminar de los planes de gobierno las palabras que no estuvieran en los textos de referencia (ODS o pactos PND), para luego vectorizarlos fácilmente como nuevas filas en la misma matriz (ODS o PND según corresponda). En suma, se tienen dos bolsas de palabras, una para ODS y otra para los pactos del PND. En ambas matrices, tanto las cadenas de texto de los planes territoriales, como los textos de referencia, corresponden a un vector fila construido de tal manera que se garantice el mismo espacio vectorial.

Posteriormente, para cada plan de gobierno se procedió a calcular la similitud coseno entre el vector “plan de gobierno” y cada uno de los vectores ODS y pactos PND, obteniendo 16 medidas de similitud para cada plan de gobierno (una con respecto a cada ODS) y 25 medidas de similitud para cada pacto del PND (una con respecto a cada pacto). Esta medida de similitud se escogió porque ha sido ampliamente utilizada en análisis de texto y se calcula con un bajo costo computacional, criterio de gran importancia considerando que el algoritmo final podría ejecutarse para la totalidad de los municipios y departamentos de Colombia. Finalmente, se calculó el promedio de las 16 medidas para tener una medida agregada, lo que corresponde a un indicador de alineación con todos los ODS que, aunque no tiene un

<sup>1</sup> Disponibles públicamente en el enlace: <https://www.un.org/sustainabledevelopment/es/objetivos-de-desarrollo-sostenible/>

<sup>2</sup> Disponibles públicamente en el enlace: <https://www.un.org/sustainabledevelopment/es/objetivos-de-desarrollo-sostenible/>

<sup>3</sup> Informe del proyecto se encuentra disponible en:

<https://colaboracion.dnp.gov.co/CDT/Desarrollo%20Digital/Big%20Data/02%20ODS/02%20Informe.pdf>



significado matemático (como sí lo tienen las 16 medidas), tiene un gran valor a nivel práctico. Así mismo, se calculó la medida agregada correspondiente a los 25 pactos del PND.

- **Análisis de Indicadores**

Por otra parte, se incluyó dentro del análisis de similitud una aproximación basada en los indicadores ODS y PND contenidos en Planes de Inversión. En este aspecto se realizó una búsqueda de cada uno de los indicadores al interior de los planes de desarrollo territoriales. De este modo se logró identificar la proporción de indicadores de cada una de las metas ODS, al igual que la proporción de indicadores de cada pacto del PND, que son mencionadas en cada uno de los planes. En este caso se incluyó la totalidad de las metas ODS y pactos PND en el análisis debido a que todos cuentan con un listado particular de indicadores. Por otra parte, se realizó un reporte para cada plan territorial en el que se especifican los indicadores que aparecen en el documento y las páginas en las que son mencionados.

La metodología utilizada para determinar si un indicador aparece o no en una página específica de un PDT consta de dos etapas principales. La primera etapa consiste en realizar un procesamiento del texto de los indicadores y de los planes territoriales. Dicho procesamiento se realizó con R y consistió en la transformación del texto a minúsculas, la remoción de números, signos de puntuación, unidades de medida de los indicadores y demás caracteres que no fueran letras; también se removieron conectores, preposiciones y palabras que no agregan significado al texto, entre las cuales se incluyeron nombres no ambiguos de las entidades territoriales y de personas. Adicionalmente se aplicó un procedimiento de *stemming* a las palabras dentro de los textos PDT y los indicadores. Esto con el fin de generalizar las palabras a su significado básico. En este sentido, se eliminan de las palabras posibles prefijos, sufijos y morfemas. Esto hace que, por ejemplo, las palabras “señora”, “señores”, “señoras” sean reducidas en el algoritmo de búsqueda a la palabra “señor”. De este modo se busca maximizar la identificación de los indicadores dentro de los textos.

- **Consolidación de los resultados en una herramienta de visualización**

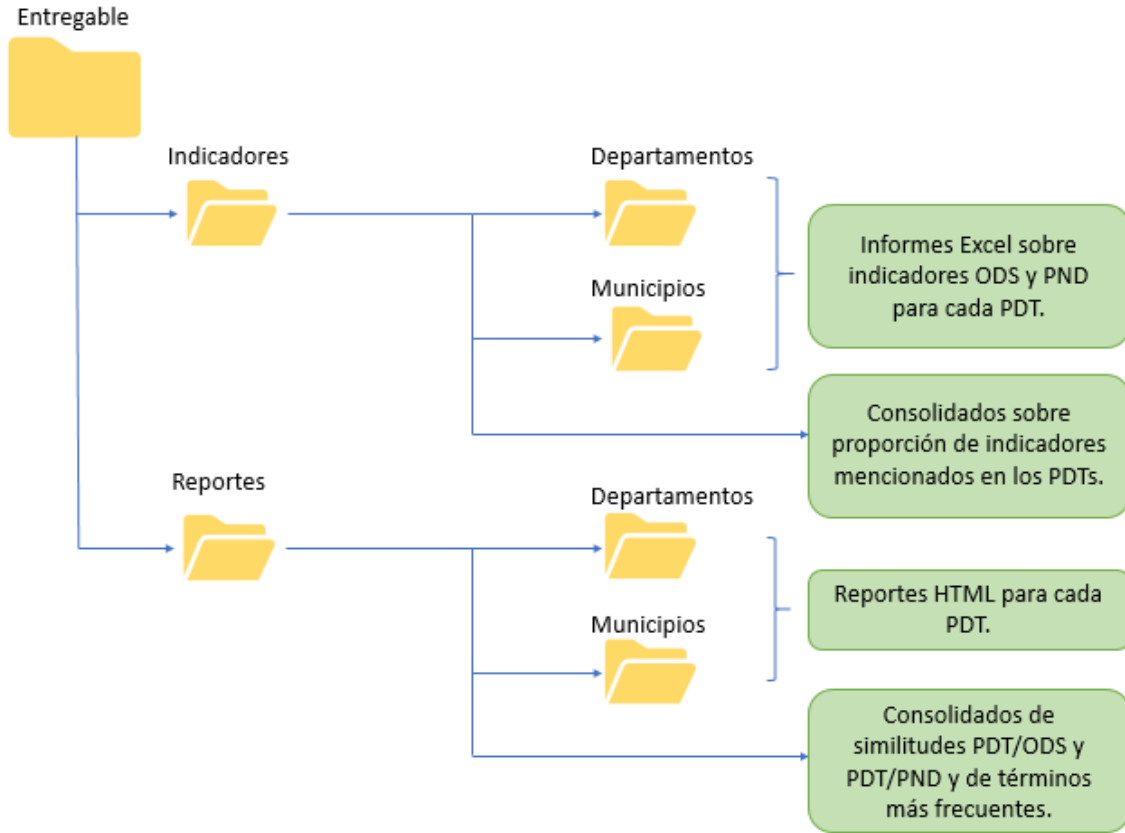
Habiendo concluido el análisis, se diseñó una plantilla en HTML utilizando RMarkdown y flexdashboard, librerías de R que facilitan la integración de distintos elementos de texto y gráficos con los cuales el usuario puede interactuar desde su navegador. En esta plantilla se incluye el resultado del análisis descriptivo y de similitud de cada uno de los planes, al igual que el resultado del análisis de indicadores ODS y PND.

## Resultados

Se realizó el procesamiento de 31 planes de desarrollo departamental y 202 planes de desarrollo municipal, según la metodología indicada. Los resultados están contenidos en una serie de carpetas con los siguientes elementos: Reportes HTML para cada uno de los planes, reportes Excel con información específica respecto a los indicadores ODS y PND encontrados en los planes y documentos con información consolidada. La Figura 2 muestra la estructura de la carpeta que contiene dichos documentos. Además, las figuras 3 y 4 muestran parte del contenido en los reportes HTML.



Figura 2: Representación de la carpeta entregada



Fuente: Elaboración propia.

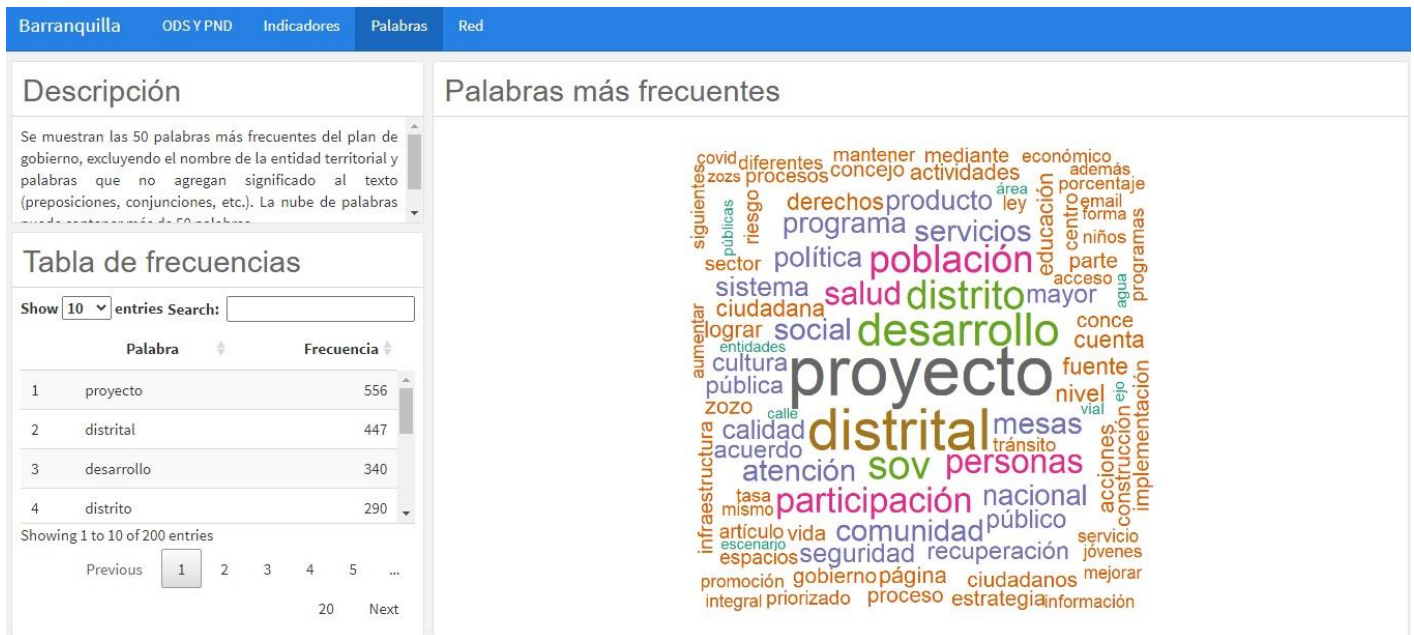
Figura 3: Vista del módulo de similitudes del reporte HTML.



Fuente: Elaboración propia



Figura 4: Vista del módulo de palabras más frecuentes del reporte HTML.



Fuente: Elaboración propia.

## Conclusiones

1. Se desarrolló una herramienta que permite determinar que tan alineados se encuentran los PDT con las estrategias de país.
2. Los análisis descriptivos realizados facilitan la identificación de temas a los cuales se da relevancia en cada Plan de Desarrollo Territorial.
3. El algoritmo desarrollado permite estimar el nivel de articulación de los planes de desarrollo con los ODS y con los pactos del PND.
4. Se implementó una aproximación basada en el listado de indicadores presentes en Planes de Inversión que podría facilitar la labor de seguimiento de los Planes de Desarrollo Territorial.
5. Mediante este proyecto se espera contribuir a un mejor acceso a fuentes de financiación por parte de las entidades territoriales.

## Socialización

El proyecto se socializó con la Dirección de Seguimiento y Evaluación de Políticas Públicas.