



DEPARTAMENTO NACIONAL DE PLANEACIÓN



DEPARTAMENTO
NACIONAL DE PLANEACIÓN

Clasificación automática de indicadores del KPT



UNIDAD DE CIENTÍFICOS DE DATOS
Dirección de Economía Naranja y Desarrollo
Digital

Diciembre, 2022

1. Introducción

2. Descripción del proyecto

3. Metodología




4. Resultados

5. Conclusiones



1. Introducción

Clasificación automática de indicadores del KPT

<h2>¿Qué?</h2>	<p>Clasificación de indicadores del KPT y asignación de sectores a productos que no se encuentran homologados o que lo están de forma errónea</p>	
<h2>¿Para qué?</h2>	<ol style="list-style-type: none">1. Facilitar el análisis sectorial2. Conocer el destino de las inversiones3. Identificar las soluciones que se plantean ante diferentes tipos de problemas	
<h2>Logros</h2>	<ol style="list-style-type: none">1. Identificación de indicadores de producto homologados y no homologados2. Asignación de sector y código de indicador a indicadores de producto no homologados3. Top 10 indicadores de bienestar por sector4. API para clasificación automática de textos y matrices nuevas	

2. Descripción del proyecto

Descripción del proyecto

Objetivo



Clasificar los indicadores de producto y bienestar del Kit de Planeación Territorial tomando como referencia las bases de datos de indicadores de bienestar de la Dirección de Seguimiento y Evaluación de Políticas Públicas, Dirección de Inversiones y Finanzas Públicas y Terridata mediante algoritmos de machine learning que permitan asignar un sector a aquellos indicadores que no se encuentran homologados o que lo están de forma errónea.

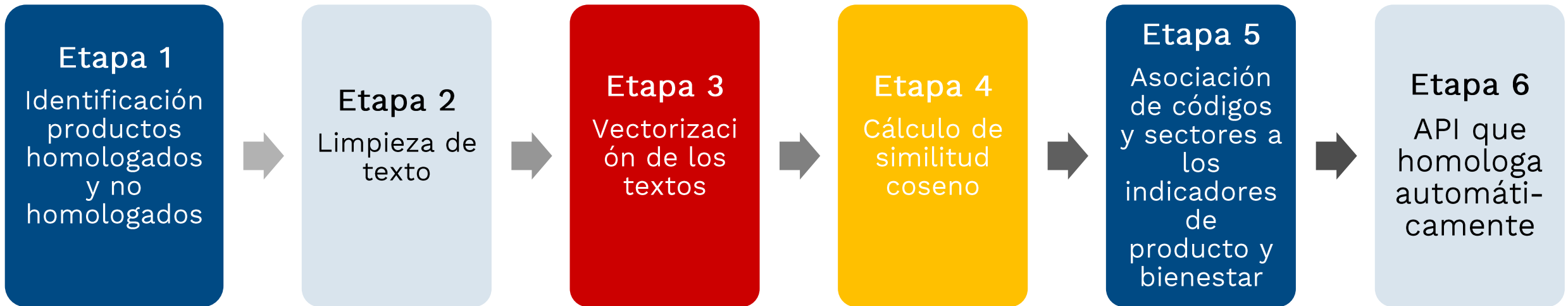
- Kit de Planeación Territorial
- Indicadores sectoriales y étnicos, que conforman el PND 2018-2022

Fuentes de
datos



3. Metodología

Metodología general



Metodología general

Etapa 1
Identificación
productos
homologados
y no
homologados



- 1
- 2
- 3

Código indicador
de producto

401223456

12|201675

12|201654

Código indicador de
producto homologado

20146970907



Metodología general

Etapa 2
Limpieza de
texto



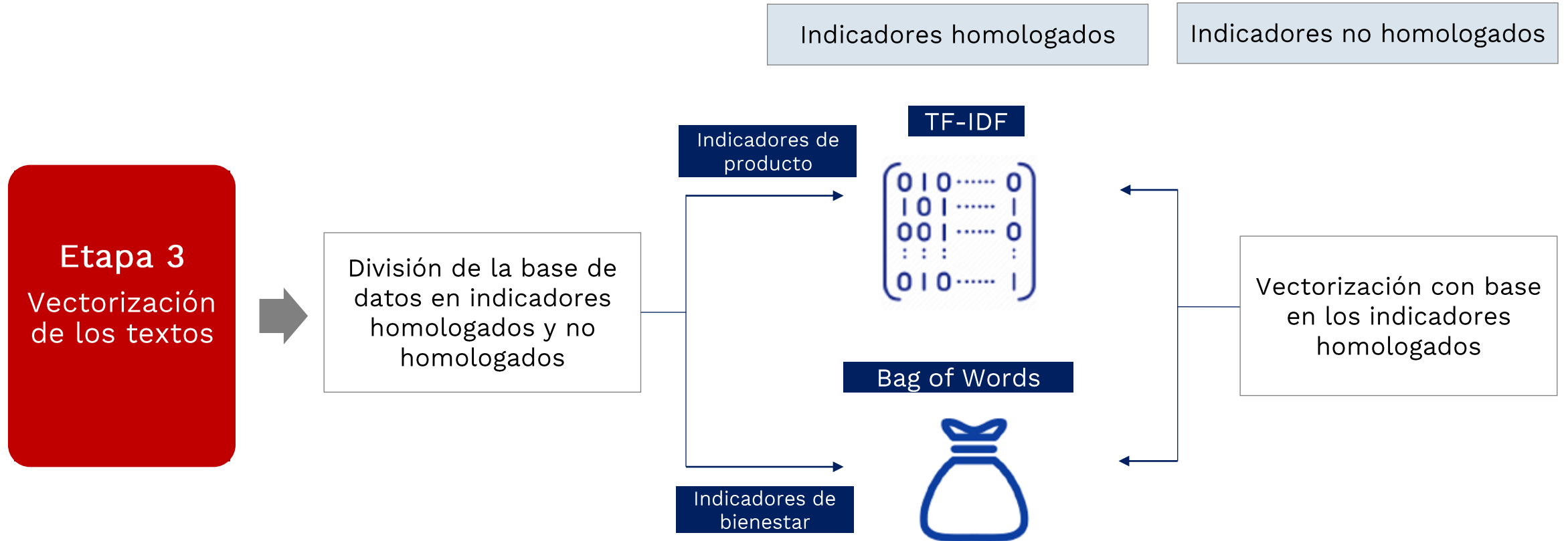
Remoción de signos de puntuación y números, pasar a minúsculas

Remoción de stopwords

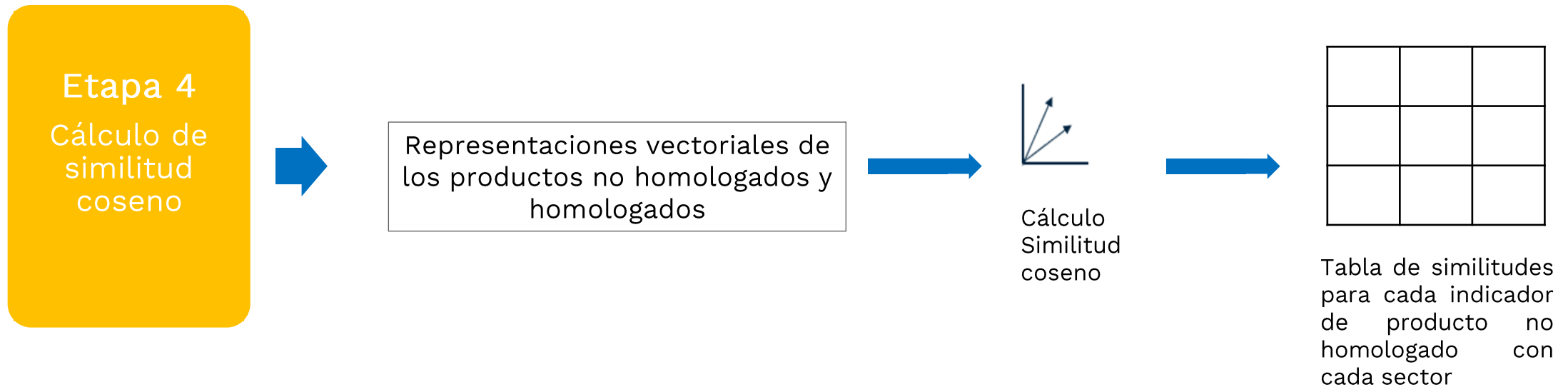
Lematización del texto

Metodología general

Indicadores de producto



Metodología general



Metodología general

Indicadores de producto

Etapa 5

Asociación de
sectores a
los
indicadores
no
homologados

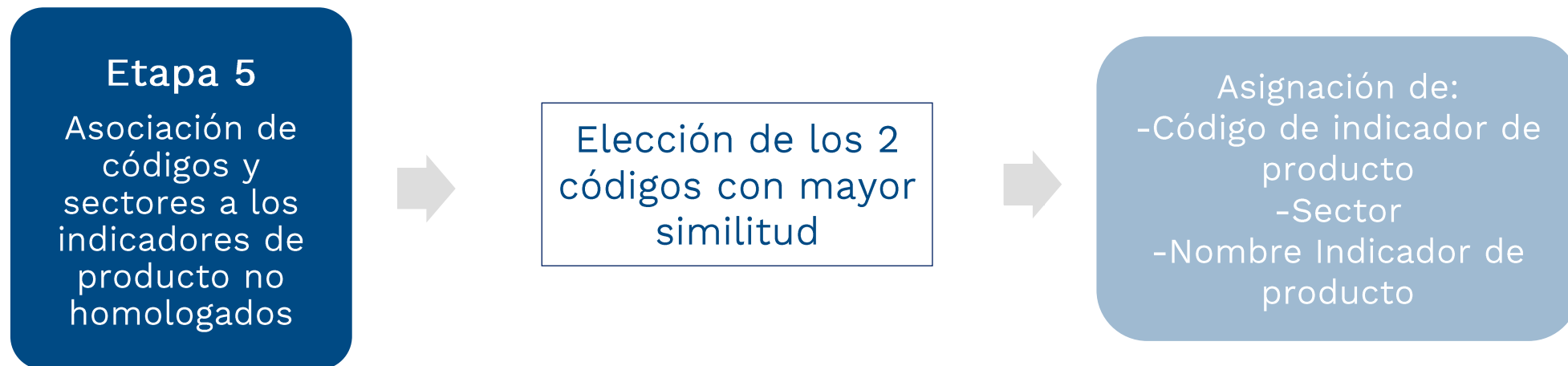


Elección de los 2
sectores con
mayor similitud



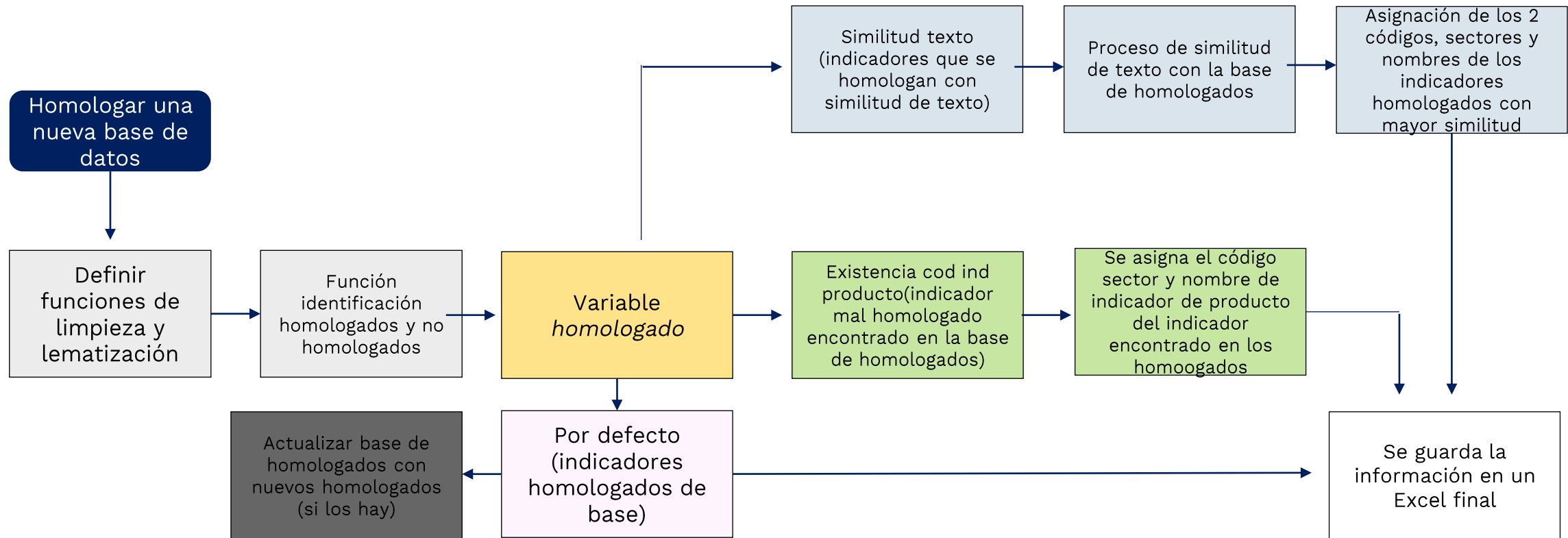
Metodología general

Indicadores de bienestar



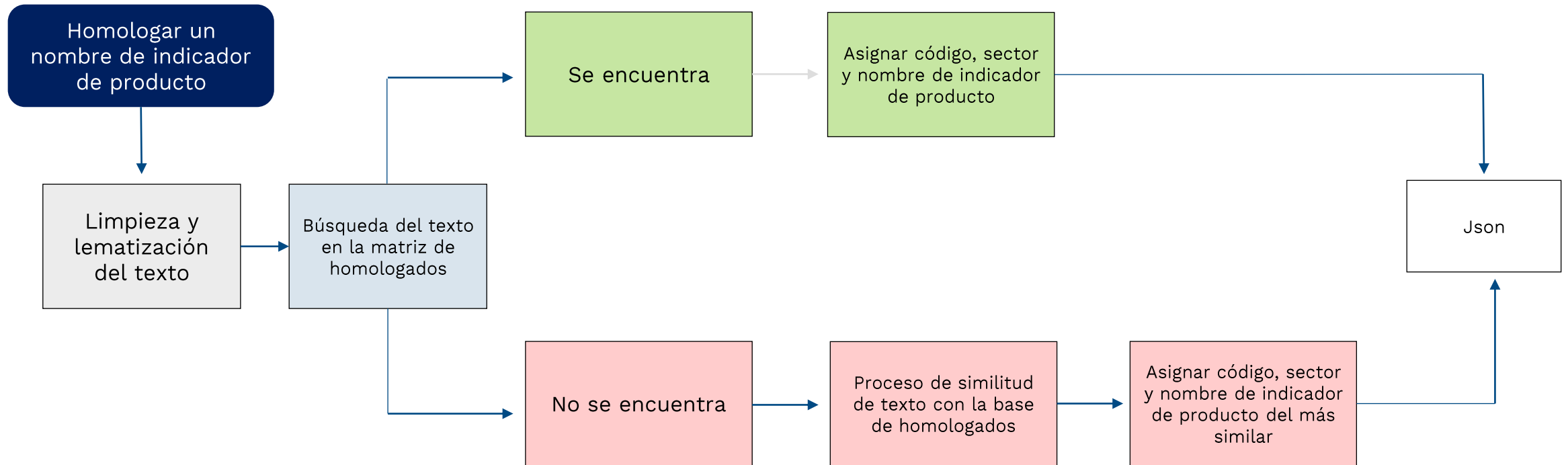
Metodología general

API



Metodología general

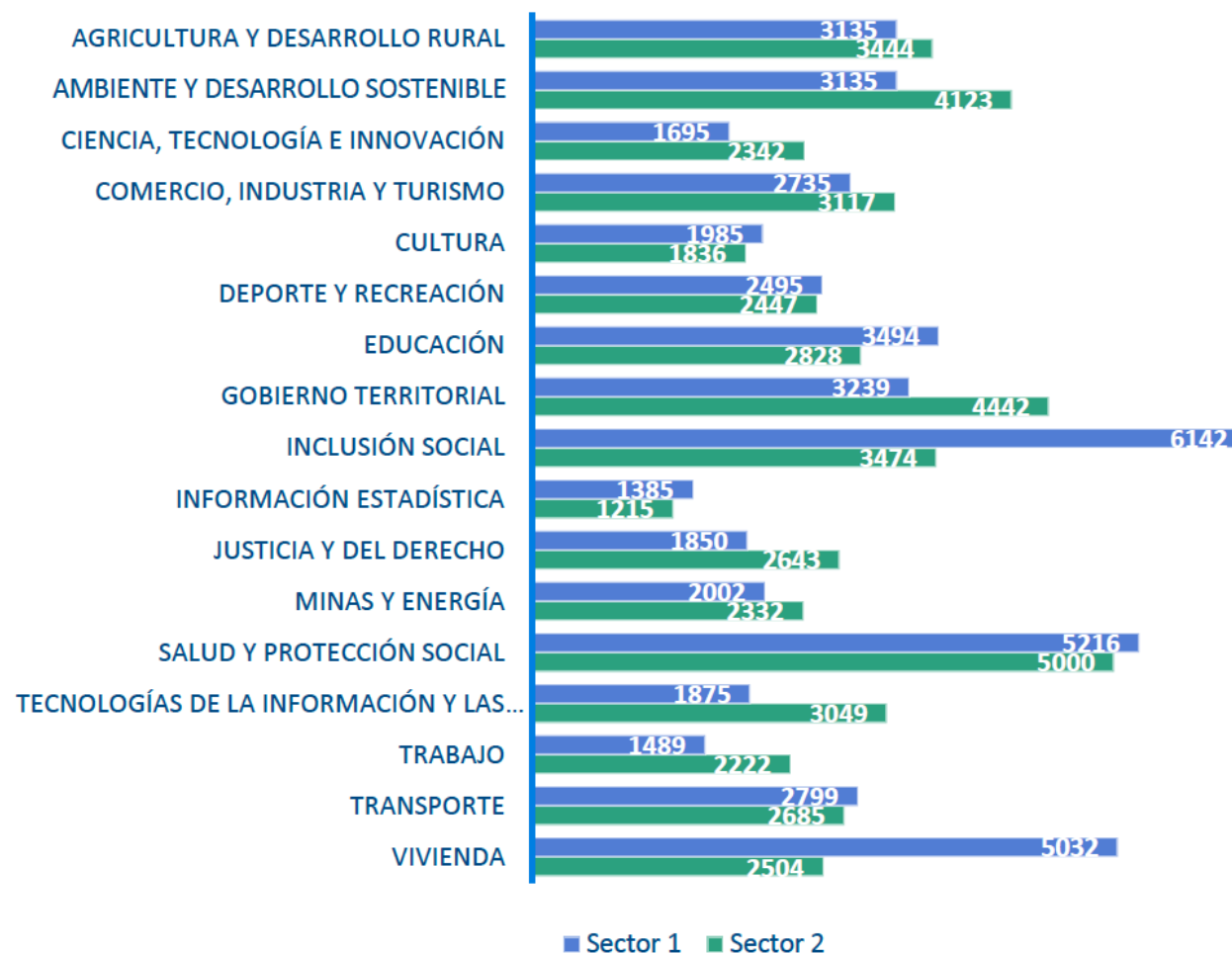
API



4. Resultados

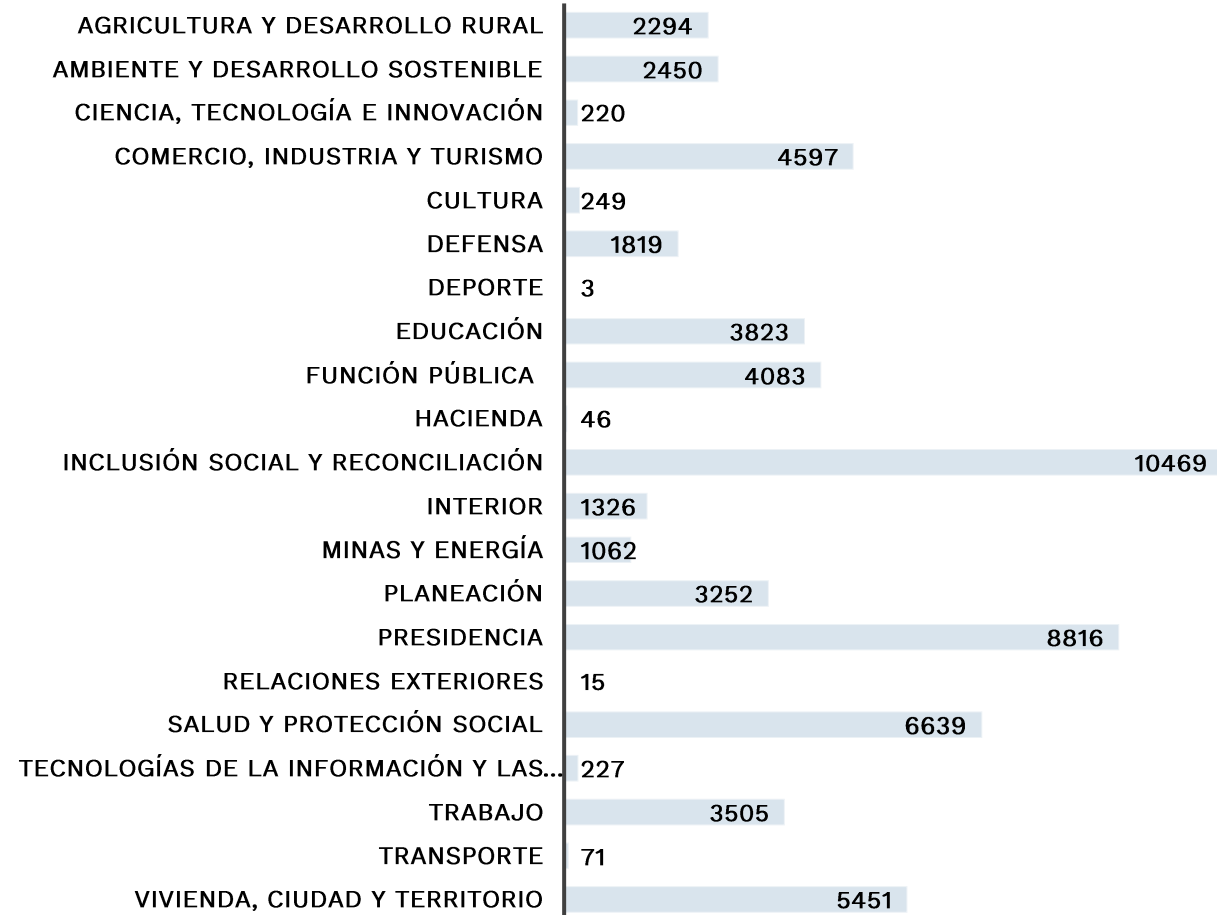
Resultados

Indicadores de producto



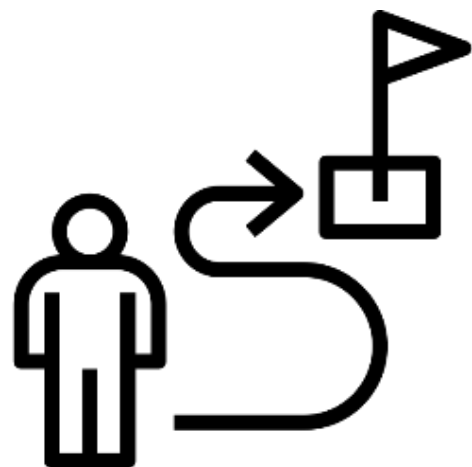
Resultados

Indicadores de bienestar



5. Conclusiones

Conclusiones



1

El proceso de identificación de los indicadores de bienestar y producto del KPT facilita que las entidades puedan incluir nuevos indicadores y sean clasificados de forma adecuada con base en los que ya están documentados a través de técnicas de similitud y análisis de texto.

2

De los nuevos indicadores homologados el sector con mayor cantidad es inclusión social y reconciliación.

3

La API desarrollada permite la homologación de textos y nuevas matrices de datos



DEPARTAMENTO NACIONAL DE PLANEACIÓN