



**El futuro  
es de todos**

**DNP**  
Departamento  
Nacional de Planeación



El futuro  
es de todos

DNP  
Departamento  
Nacional de Planeación




# Rastreo de proyectos de inversión relacionados con Cambio Climático (Actualización y mejora)

Unidad de Científicos de Datos  
Dirección de Desarrollo Digital

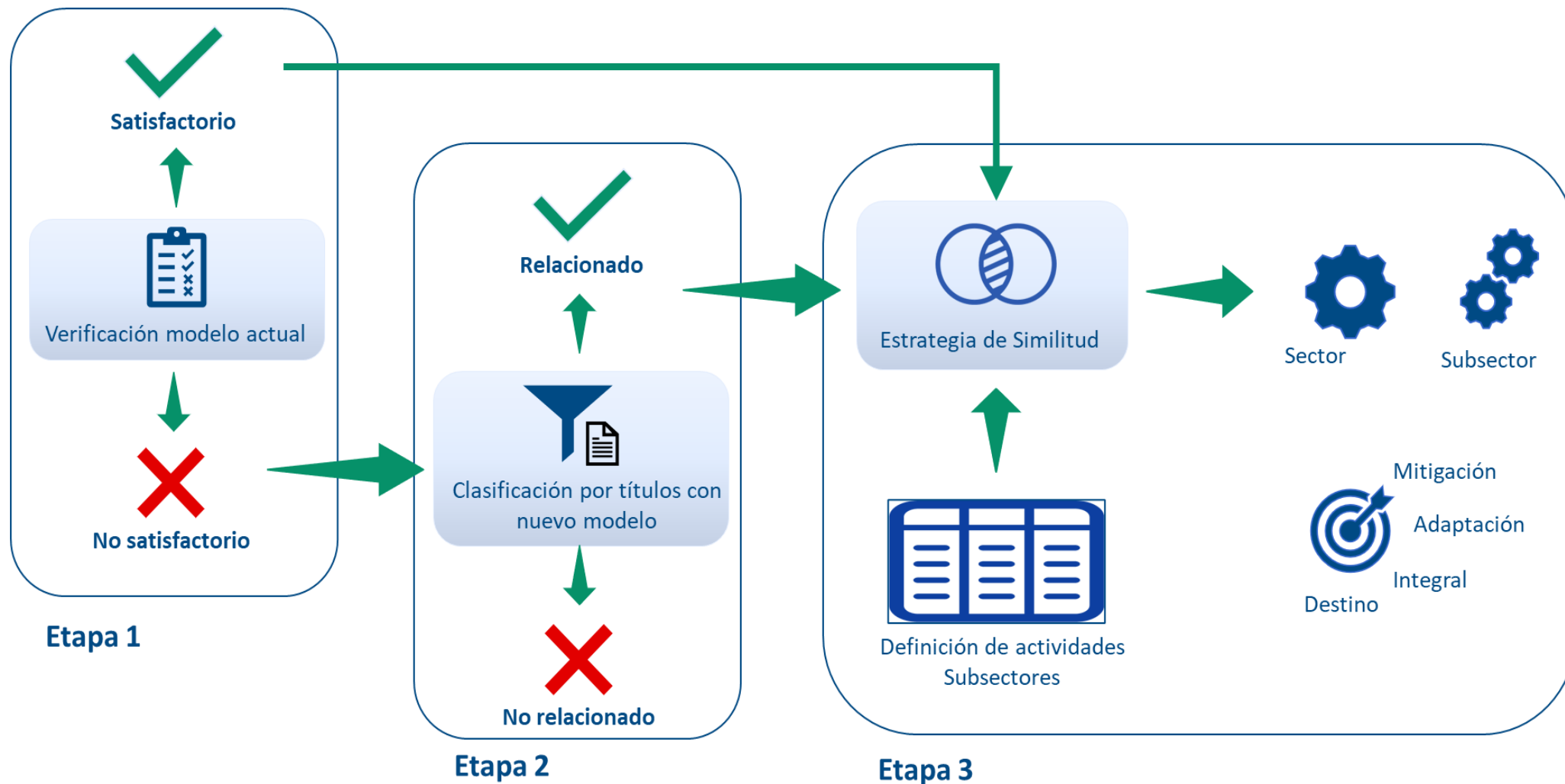
Abril, 2021



# Rastreo de proyectos de inversión relacionados con Cambio Climático

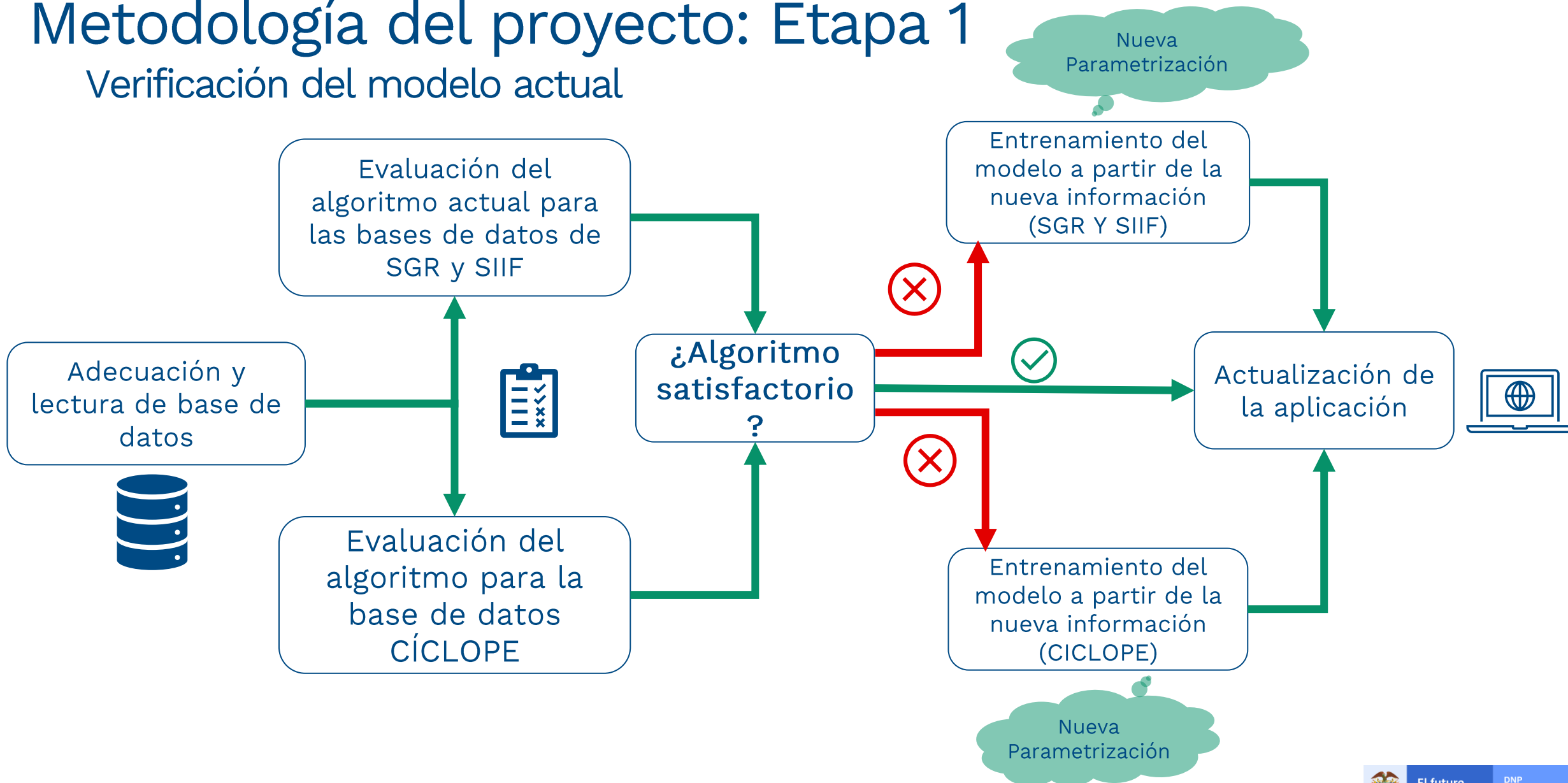
<b>¿Qué?</b>	Fortalecer y adecuar el algoritmo de rastreo de proyectos de inversión relacionados con cambio climático en las bases de datos SGR, SIIF y CÍCLOPE	
<b>¿Para qué?</b>	Facilitar la tarea de rastreo de proyectos de inversión realizado por el equipo de la DADS de manera semi-manual	
<b>Resultados</b>	<ul style="list-style-type: none"><li>- Actualización modelo Machine Learning de rastreo de proyectos de inversión de SGR – SIIF y Cíclope</li><li>- Actualización de herramienta de visualización de resultados</li></ul>	

# Metodología del proyecto



# Metodología del proyecto: Etapa 1

## Verificación del modelo actual



# Metodología del proyecto: Etapa 2

## Algoritmo de rastreo

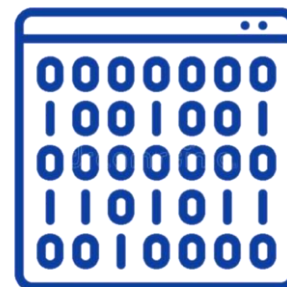


Títulos de los proyectos



Limpieza de texto

Corrección ortográfica,  
quitar tildes, espacios  
extra, números, etc.



TF - IDF

Unigramas  
Bigramas



Modelo de  
clasificación

SVM



# Metodología del proyecto: Etapa 3

## Estrategia de similitud: Taxonomía

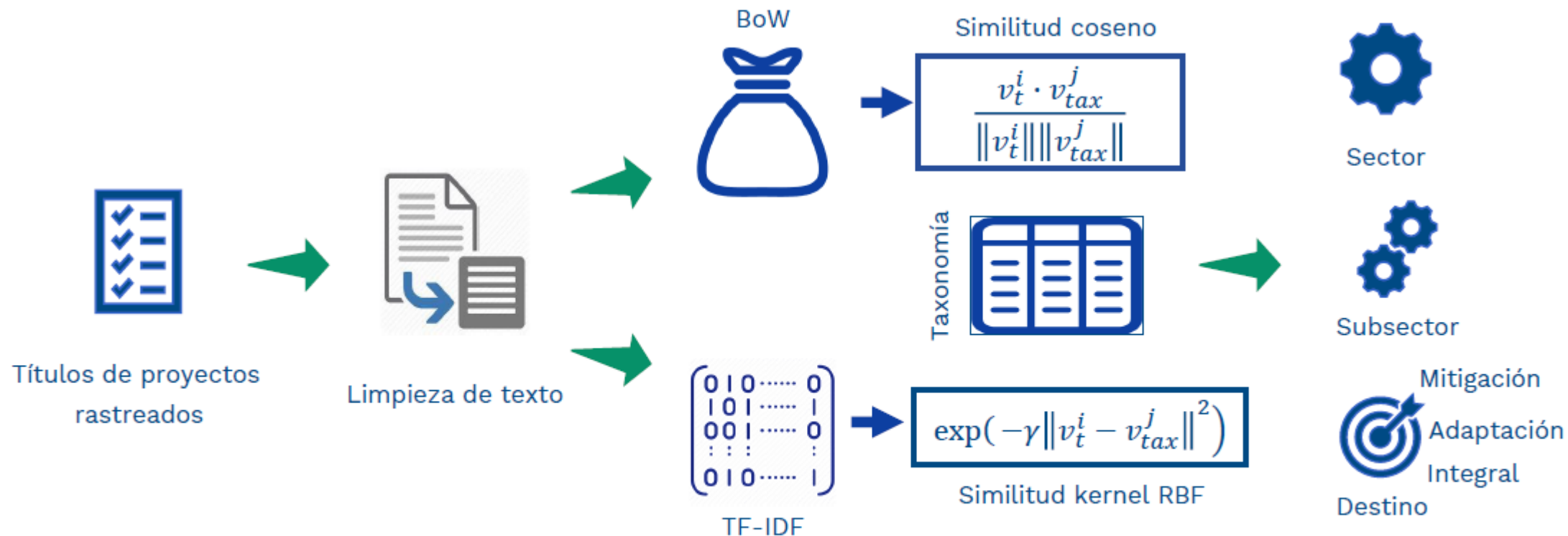
Sector	Subsector	Actividades	M	A	I
Energía	Generación, mejora y acceso de electricidad	Generar energía con fuentes no convencionales en zonas no interconectadas (sistemas híbridos)	1		
	Generación, mejora y acceso de electricidad	Acceder a la energía a través de la electrificación rural		1	
	Eficiencia energética	Usar energía solar para calentamiento de agua	1		
	Producción minera	Investigar y desarrollar capacidades para mejorar la resiliencia de la actividad minera		1	
Agropecuario	Desarrollo rural	Promover sistemas agroforestales			1
	Agricultura	Usar los residuos de cosecha para la generación de energía	1		

ID	Sector	Subsectores	Actividades
1	Agropecuario	4	37
2	Educación	2	3
3	Energía	5	34
4	Gestión del riesgo y atención de desastres	1	8
5	Industria	7	32
6	Medio Ambiente y Recursos Naturales	3	43
7	Residuos	1	15
8	Salud	2	3
9	Transporte	3	40
10	Transversal	3	24
11	Turismo	2	6
12	Vivienda	2	4
<b>Total</b>		<b>35</b>	<b>249</b>



# Metodología del proyecto: Etapa 3

Estrategia de similitud: Título – Descripción de actividad

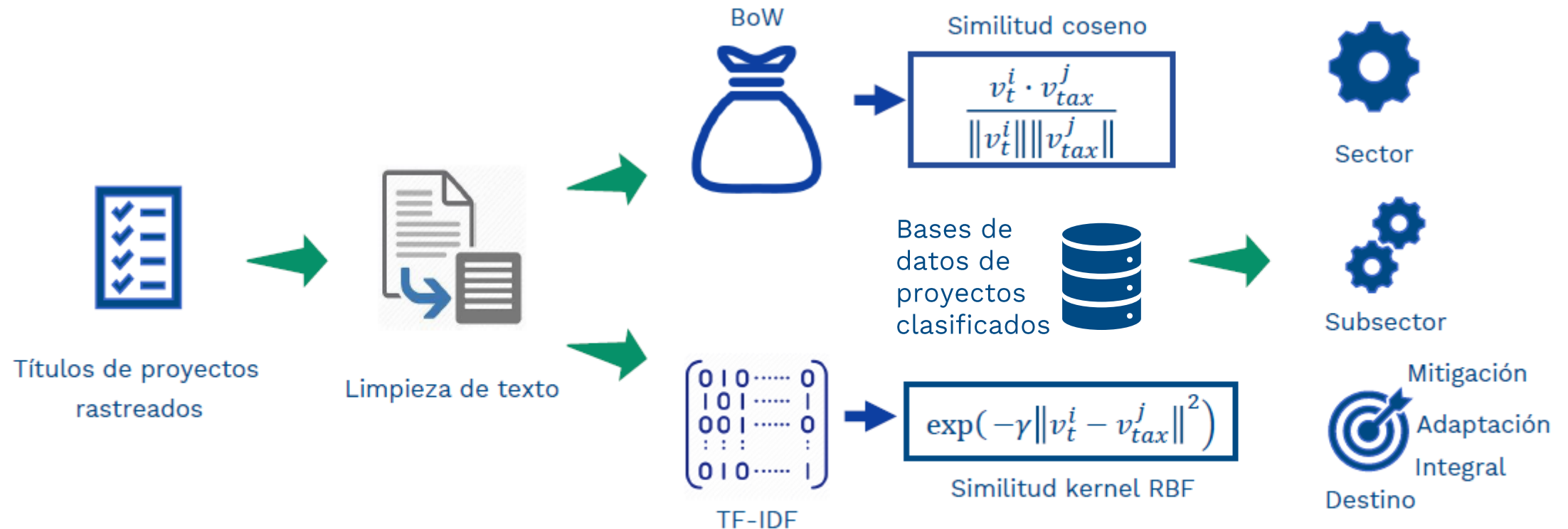


Donde  $v_t^i$  es el vector numérico del título  $i$ ,  $v_{tax}^j$  es el vector numérico del texto de la actividad  $j$ .



# Metodología del proyecto: Etapa 3

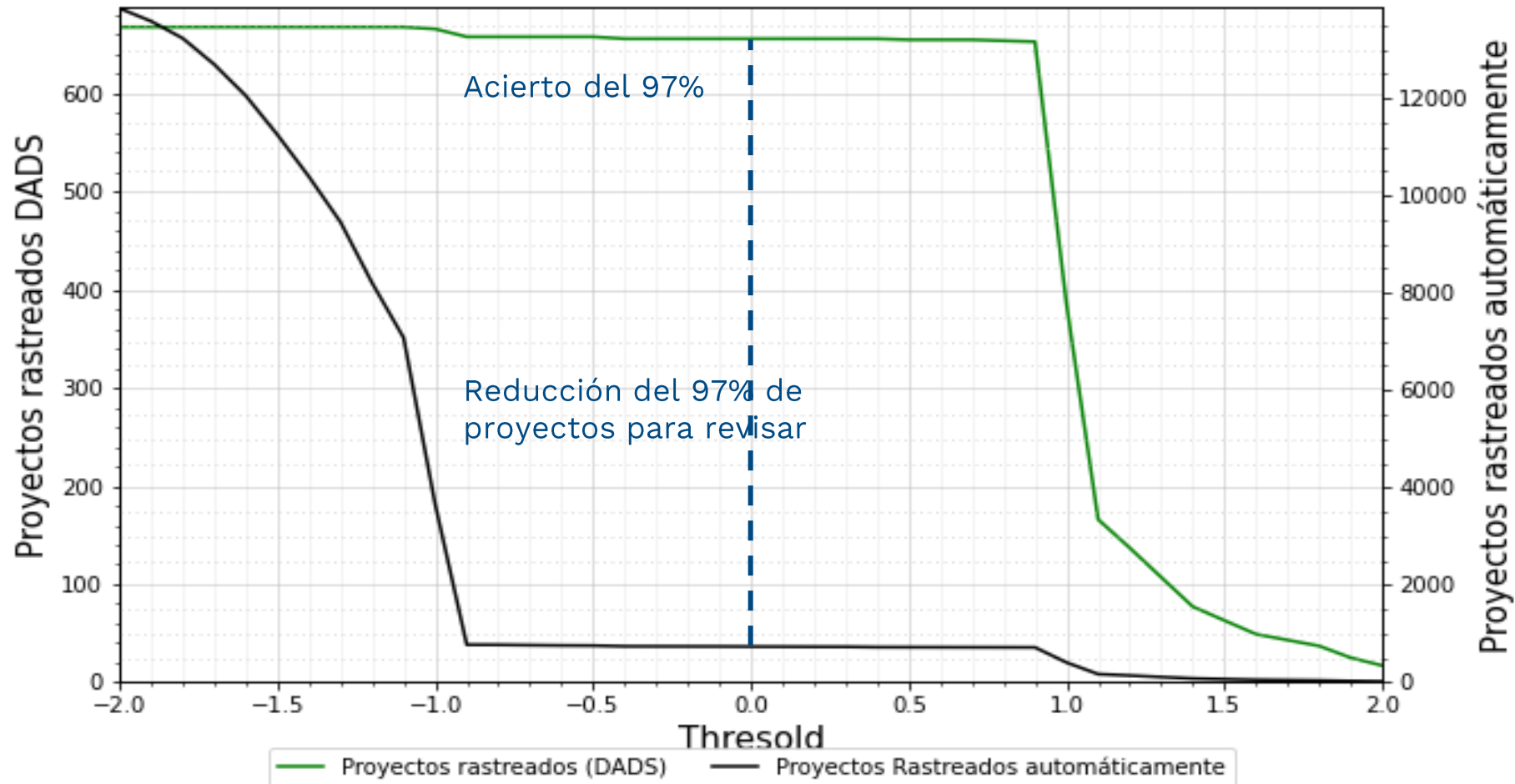
## Estrategia de similitud: Títulos



Donde  $v_t^i$  es el vector numérico del título  $i$ ,  $v_{tax}^j$  es el vector numérico del texto de la actividad  $j$ .

# Resultados

Rendimiento para sistemas de información SGR y SIIF (modelo 2021)



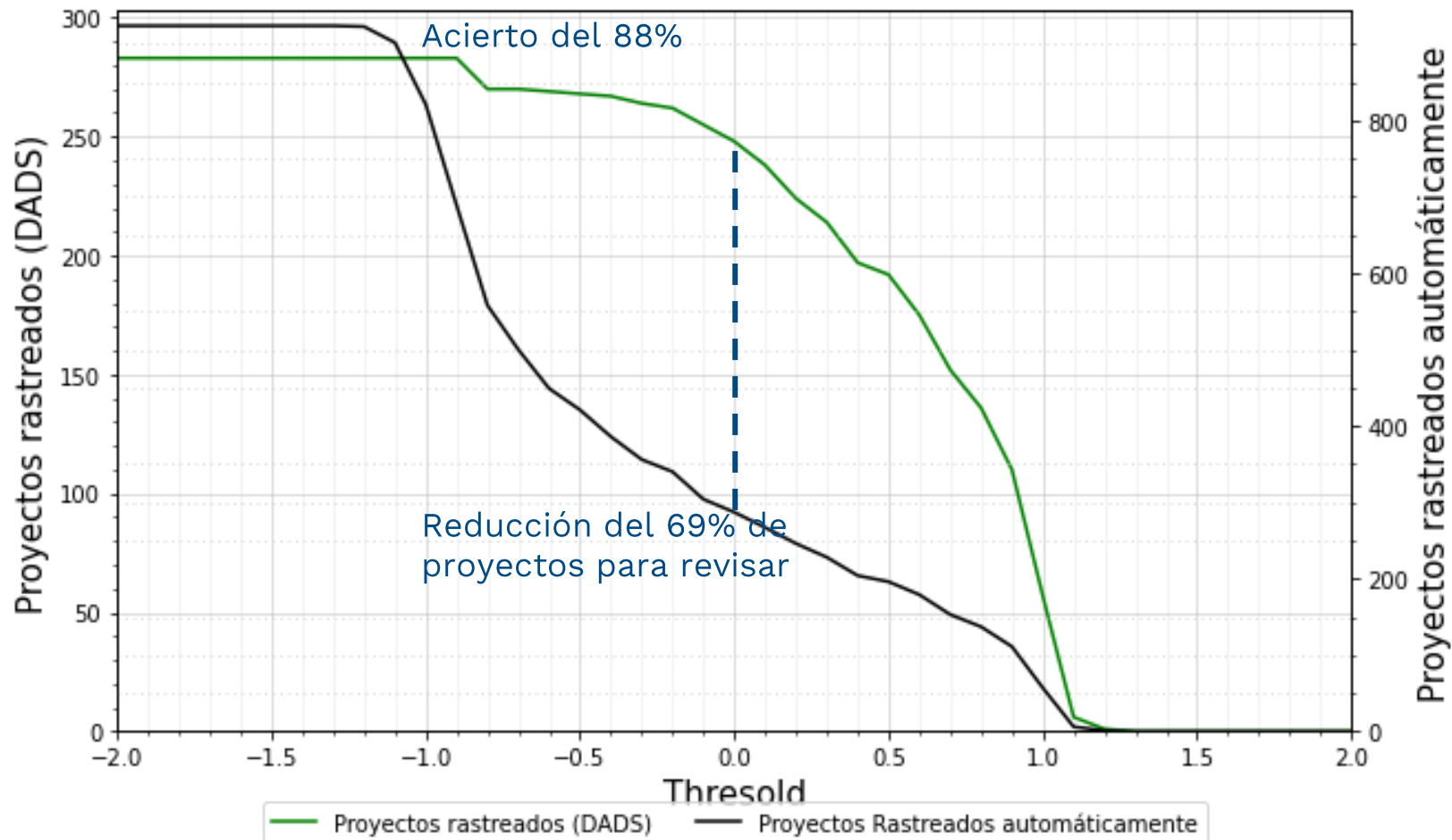
Modelo 2020:

- Reducción de proyectos en un 83%
- Acierto 90%



# Resultados

Rendimiento para sistema de información Cíclope (modelo 2021)



**Modelo 2020:**

- Reducción de proyectos en un 25%
- Acierto 90%

# Resultados

## Estrategias de similitud M1

### Resultados SGR - SIIF

Sectores	Precisio n	Recall	F1 - score	Muestras
Agropecuario	0.51	0.48	0.49	98
Educación	0.17	0.23	0.19	13
Energía	0.26	0.49	0.34	84
Gestión del riesgo y atención de desastres	0.75	0.30	0.43	132
Industria	0.04	0.20	0.07	5
Medio Ambiente y Recursos Naturales	0.64	0.67	0.66	422
Residuos	0.71	0.26	0.38	78
Salud	0.00	0.00	0.00	0
Transporte	0.74	0.89	0.01	54
Transversal	0.38	0.22	0.28	92
Turismo	0.00	0.00	0.00	1
Vivienda	0.00	0.00	0.00	0
Promedio	0.59	0.51	0.52	979

### Resultados CICLOPE

Sectores	Precisio n	Recall	F1 - score	Muestras
Agropecuario	0.43	0.39	0.41	38
Educación	0.00	0.00	0.00	6
Energía	0.64	0.35	0.45	26
Gestión del riesgo y atención de desastres	0.60	0.33	0.43	18
Industria	0.19	0.75	0.30	4
Medio Ambiente y Recursos Naturales	0.56	0.61	0.58	84
Residuos	0.50	0.75	0.60	4
Salud	0.00	0.00	0.00	0
Transporte	0.39	0.80	0.52	15
Transversal	0.42	0.32	0.36	84
Turismo	0.08	1.00	0.15	1
Vivienda	0.25	0.33	0.29	3
Promedio	0.48	0.45	0.45	283

# Resultados

## Estrategias de similitud M2

### Resultados SGR - SIIF

Sectores	Precisio n	Recall	F1 - score	Muestras
Agropecuario	0.51	0.52	0.52	98
Educación	0.38	0.46	0.41	13
Energía	0.36	0.36	0.36	84
Gestión del riesgo y atención de desastres	0.75	0.38	0.50	132
Industria	0.00	0.00	0.00	5
Medio Ambiente y Recursos Naturales	0.69	0.68	0.69	422
Residuos	0.61	0.26	0.36	78
Salud	0.00	0.00	0.00	0
Transporte	0.69	0.81	0.75	54
Transversal	0.38	0.26	0.31	92
Turismo	0.00	0.00	0.00	1
Vivienda	0.00	0.00	0.00	0
<b>Promedio</b>	<b>0.61</b>	<b>0.52</b>	<b>0.55</b>	<b>979</b>

### Resultados CICLOPE

Sectores	Precisio n	Recall	F1 - score	Muestras
Agropecuario	0.46	0.34	0.39	38
Educación	0.00	0.00	0.00	6
Energía	0.67	0.54	0.60	26
Gestión del riesgo y atención de desastres	0.57	0.44	0.50	18
Industria	0.23	0.75	0.35	4
Medio Ambiente y Recursos Naturales	0.64	0.63	0.63	84
Residuos	0.14	0.25	0.18	4
Salud	0.00	0.00	0.00	0
Transporte	0.33	0.80	0.47	15
Transversal	0.47	0.35	0.40	84
Turismo	0.17	1.00	0.29	1
Vivienda	0.12	0.33	0.18	3
<b>Promedio</b>	<b>0.51</b>	<b>0.48</b>	<b>0.48</b>	<b>283</b>

# Resultados

## Estrategias de similitud M3

### Resultados SGR - SIIF

Sectores	Precisio n	Recall	F1 - score	Muestras
Agropecuario	0.77	0.75	0.76	32
Educación	0.60	0.75	0.67	4
Energía	0.93	0.90	0.92	31
Gestión del riesgo y atención de desastres	0.89	0.96	0.93	71
Industria	0.00	0.00	0.00	1
Medio Ambiente y Recursos Naturales	0.86	0.89	0.88	118
Residuos	0.58	0.58	0.58	19
Salud	0.00	0.00	0.00	0
Transporte	0.81	0.94	0.87	18
Transversal	0.92	0.68	0.78	34
Turismo	0.00	0.00	0.00	1
Vivienda	1.00	1.00	1.00	1
<b>Promedio</b>	<b>0.84</b>	<b>0.84</b>	<b>0.84</b>	<b>330</b>

### Resultados CICLOPE

Sectores	Precisio n	Recall	F1 - score	Muestras
Agropecuario	0.50	0.38	0.43	8
Educación	0.00	0.00	0.00	1
Energía	0.67	0.40	0.50	5
Gestión del riesgo y atención de desastres	0.50	0.50	0.50	4
Industria	0.00	0.00	0.00	1
Medio Ambiente y Recursos Naturales	0.65	0.76	0.70	17
Residuos	0.00	0.00	0.00	1
Salud	0.00	0.00	0.00	0.00
Transporte	0.00	0.00	0.00	0.00
Transversal	0.67	0.67	0.67	3
Turismo	0.41	0.41	0.41	17
Vivienda	0.00	0.00	0.00	0.00
<b>Promedio</b>	<b>0.52</b>	<b>0.51</b>	<b>0.51</b>	<b>57</b>

# Resultados

Herramienta de visualización

Opción de selección de modelo de clasificación según el sistema de información

Límite de decisión recomendado para cada modelo

Rastreo de proyectos de inversión relacionados con cambio climático

Abrir archivo

Drag and drop file here  
Limit 200MB per file · XLSX

Browse files

Libro1.xlsx  
17.3KB

Sistema de información:  
SGR o SIIF (Modelo Anterior)

Similitud por:  
 Descripción  
 Descripción + subsector  
 Títulos

Threshold  
-2.00 -0.20 1.00

Selección sector  
Todos

El futuro es de todos  
DNP  
Departamento Nacional de Planeación

Columnas  
 BPIN x NOMBRE DEL PRO... x Score x Sector1 x Subsector1 x M1 x A1 x I1 x similitud1 x Sector2 x Subsector2 x  
 M2 x A2 x I2 x similitud2 x

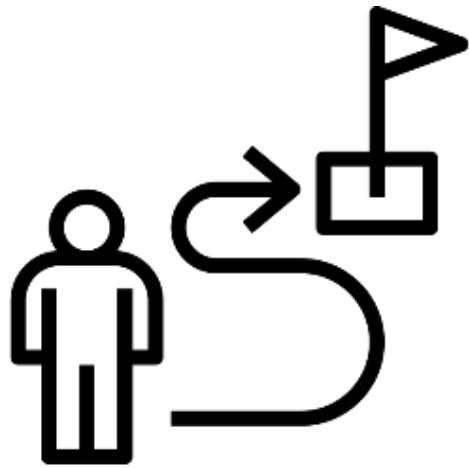
BPIN	NOMBRE DEL PROYECTO	Score	Sector1	Subsector1	M1	A1	I1	similitud1
2017000050063	RECUPERACIÓN HIDRAULICA Y...	2.6495	Energía	Producción de hidrocarbur...	1	nan	nan	0.4472
2018000050048	RECUPERACIÓN HIDRAULICA Y...	2.3568	Medio Ambiente y Recursos...	Biodiversidad y gobernanz...	nan	nan	1	0.3162
2018200450012	MEJORAMIENTO AMBIENTAL DE...	2.0130	Agropecuario	Agricultura	1	nan	nan	0.2182
2018479800007	Recuperación hidráulica y...	2.0000	Energía	Producción de hidrocarbur...	1	nan	nan	0.2582
2018475510004	RECUPERACIÓN HIDRAULICA Y...	1.7405	Energía	Producción de hidrocarbur...	1	nan	nan	0.2582
2017000070072	FORMULACIÓN DE LOS PLANES...	1.7305	Educación	Educación sobre cambio cl...			1	0.5385
2017000030138	IMPLEMENTACIÓN DE SISTEMA...	1.6521	Transversal	Producción y consumo sust...	nan		1	0.3000
2018000070015	CONSTRUCCIÓN OBRAS DE PRO...	1.6513	Transversal	Investigación y fortaleci...			1	0.5270
2017000050070	CONSTRUCCIÓN DE OBRAS DE ...	1.6492	Agropecuario	Desarrollo rural	nan	1	nan	0.3397
2018002200053	CONSERVACIÓN DE ÁREAS DE ...	1.5708	Medio Ambiente y Recursos...	Biodiversidad y gobernanz...			1	0.2182
2018000060031	CONSTRUCCIÓN DE OBRAS DE ...	1.5704	Transversal	Investigación y fortaleci...	nan	1	nan	0.3750

110 proyectos rastreados

[Descargar](#)



# Conclusiones y recomendaciones



- 1** Se presentaron mejoras en los rendimientos de los nuevos modelos supervisados, logrando así cumplir con los objetivos del proyecto
- 2** La alimentación frecuente de las bases de datos con los títulos de los proyectos clasificados es fundamental para mejoras futuras tanto en el modelo de clasificación como en el modelo de similitud
- 3** La herramienta de visualización permitirá a los usuarios de la DADS reducir el número de proyectos por revisar gracias a los modelos de clasificación para bases de datos SGR, SIIF y Cíclope





**El futuro  
es de todos**

**DNP**  
Departamento  
Nacional de Planeación