**ANEXO 1. FOMULARIO DE POSTULACIÓN DE RETOS**

Fecha: Mayo 18 de 2020

Ciudad: Bogotá

Señores

Talento Digital

Ministerio MinTIC

Atendiendo a la convocatoria realizada por ustedes, a continuación nos permitimos proponer un reto para el proceso de formación que se realizará en Ciencia de Datos durante el 2020, el cual atiende a una necesidad particular de la entidad.

Hemos leído los términos y condiciones de la convocatoria y disponemos de la información necesaria para que el equipo pueda desarrollar el proyecto.

|  |  |
| --- | --- |
| **DATOS BÁSICOS** | |
| * + - 1. Nombre de la entidad o empresa: | **Ministerio de las Tecnologías de la Información y las Comunicaciones - MinTIC** |
| 1. Tipo de empresa o entidad: | Empresa privada: \_\_\_\_\_  Sector: \_\_\_\_\_Telecomunicaciones\_\_\_\_\_\_\_\_\_\_\_  Entidad pública: \_\_\_X\_\_\_  Nacional \_\_\_\_\_\_  Territorial \_\_\_\_\_\_  Otra - Cual? \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |
| 1. Área/ departamento o Dirección de la empresa o entidad a la que está asociada el reto. | **Oficina Asesora de Planeación y Estudios Sectoriales** |
| 1. Punto de contacto operativo para el Proyecto **(Líder)** | |
| Nombres y apellidos | **Lady Velásquez Henao** |
| Cargo | **Jefe de Oficina** |
| Correo electrónico | **lvelasquezh@mintic.gov.co** |
| 1. Un punto de **contacto técnico** para el Proyecto que puede responder preguntas técnicas sobre el conjunto de datos y el problema | |
| Nombres y apellidos | **Darwin Fonseca** |
| Cargo | **Contratista** |
| Correo electrónico | **dfonseca@mintic.gov.co** |
| **DESCRIPCIÓN DEL PROYECTO** | |
| **Descripción general de la empresa o entidad pública.** | El Ministerio de Tecnologías de la Información y las Comunicaciones, según la Ley 1341 o Ley de TIC, es la entidad que se encarga de diseñar, adoptar y promover las políticas, planes, programas y proyectos del sector de las Tecnologías de la Información y las Comunicaciones.  Dentro de sus funciones está incrementar y facilitar el acceso de todos los habitantes del territorio nacional a las Tecnologías de la Información y las Comunicaciones y a sus beneficios [[1]](#footnote-2). |
| **Descripción del problema que debe resolverse** | En el Ministerio no se están aprovechando para fines estadísticos y analítica muchos de los registros administrativos [[2]](#footnote-3). Las áreas misionales de la entidad generan estos conjuntos de datos cada vez que implementa una política, plan, programa, proyecto o iniciativa donde se busca beneficiar un número de ciudadanos, empresas o municipios, dichos registros contienen información valiosa que le permite producir análisis y estadísticas oficiales a la entidad, pero en muchos casos no se está utilizando, pese a la información de miles de beneficiarios consignadas en dichas bases, que permitirían entre otras cosas; mejorar los programas, hacer seguimiento a la efectividad del mismo, informar a los ciudadanos de la gestión del Ministerio etc.  Desde la Oficina Asesora de Planeación y Estudios Sectoriales se inició un ejercicio con los registros administrativos de un área de la entidad que dispuso de ellos para su análisis y posterior visualización de los resultados de esa gestión. Sin embargo, al recibir el conjunto de datos se realizó un ejercicio de revisión y limpieza de esta, identificando aspectos en la calidad de los datos (completitud y estandarización de las variables), lo anterior afecta el análisis de las bases suministradas y por ende la elaboración de análisis e indicadores que de allí se derivan, lo que nos obligó a validar con el área para garantizar la publicación de información amplia, exacta y veraz, cumpliendo con el principio de claridad y exactitud.  En la labor de revisión y limpieza de la data se identificaron:   * Se presentan registros duplicados para la variable. * Se presentan registros vacíos en la variable. * Se presentan registros sin diligenciamiento en la variable * La variable Teléfono, por ejemplo, no se encuentra estandarizada hay diferentes tipos de teléfono fijo, celular y al diligenciar aparentemente permite el ingreso de texto, adicionalmente los valores en la variable varían ente 7 y 14 dígitos. * La variable `Contacto: E-mail` no se encuentra estandarizada ya que en los registros se observan caracteres especiales con extensión de correo, números de celular con extensiones de correo.   Una vez realizada esa labor se entrará al análisis de la información, la construcción de indicadores y la visualización de estos.  Lo anterior se debe realizar con cada uno de los registros administrativos construido por el Ministerio a través de los años, y hacer analítica descriptiva y predictiva con ellos. |
| **¿Por qué es importante este problema?** | Porque al solucionarlo podemos Identificar y promover el aprovechamiento estadístico de registros administrativos, ampliar la producción de información estadística, hacer analítica descriptiva y predictiva, dado que se estaría identificación, fortalecimiento y gestionando datos, lo que nos permitiría alinearnos a las metodologías del DANE y a la política de Gestión y desempeño Institucional de Información Estadística.  Igualmente se tendrían datos consolidados para producir información relacionada con los beneficiarios y logros de los diferentes programas desarrollados por MinTIC y evaluar sus resultados e impacto.  Permite visualizar de forma fácil, así como la gestión del Ministerio históricamente.  Se construiría conocimiento valioso, se mejorarían procesos, y contribuiría a la toma de decisiones.  La propuesta estaría definida en dos etapas, **la primera de ellas es gestión de los datos y la analítica de datos.** En la primera etapa se debe realizar la búsqueda y recolección de las fuentes de datos en las diferentes áreas misionales del Ministerio, la identificación de variables que permitan clasificar los conjuntos de datos, explorar los datos para encontrar errores e inconsistencias, limpiar y depurar los conjuntos de datos, generando las variables que sean necesarias, aplicando procesos de imputación de datos, creando nuevas categorías de clasificación y eliminando duplicados de los datos.  **La segunda etapa, comprende la aplicación de técnicas de modelamiento y análisis**, mediante el uso de herramientas y servicios tecnológicos que permitan gestionar y procesar los datos para extraer conclusiones que permitan dar respuesta a los objetivos establecidos, los cuales serán presentados y puestos a disposición mediante el empleo de herramientas de visualización interactivas, documentos, notas estadísticas o estudios, que permitan ilustrar las conclusiones extraídas como información de calidad y oportuna, que oriente y fundamente la toma de decisiones o permitan que los grupos de interés conozcan el impacto de los programas del Ministerio en la población colombiana. |
| **¿Cuáles son los conjuntos de datos que los participantes pueden acceder para resolver este problema?** | Todos los registros administrativos y bases de datos generadas por las áreas misionales del ministerio que se han generado en los últimos 10 años (en los casos que se cuente desde ese periodo) a la fecha. Pero estos deben ser gestionados con cada una de ellas. |
| **Enumere y describa el conjunto de datos específicos** | Esta en una de las actividades que contempla este proyecto en su primera etapa, se debe realizar la búsqueda, identificación y recolección de las fuentes de datos generados por las diferentes áreas misionales del Ministerio, la identificación de variables que permitan clasificar los conjuntos de datos, explorar los datos para encontrar errores e inconsistencias, limpiar y depurar los conjuntos de datos, generando las variables que sean necesarias, aplicando procesos de imputación de datos, creando nuevas categorías de clasificación y eliminando duplicados de los datos. |
| **¿Qué tipo de conjuntos de datos externos pueden ser de interés para el problema?** | Para los análisis se podrían utilizar los generados por el DANE relacionados con la población colombiana, así como los indicadores relacionados con el sector TIC que esta entidad produce.  Información del Cubo de Pila del Ministerio de Salud.  Información del SENA y del Ministerio de Educación, por ejemplo, para evaluar si la formación en TIC impartida a través de los programas desarrollados por el ministerio como están contribuyendo a la formación Técnica de los Colombianos. |
| **Resultados esperados del proyecto.** | Los resultados esperados son:  La construcción de conocimiento valioso, la mejora de procesos, y contribuir a la toma de decisiones.  Estar alineados con lo establecido en la Política de gestión y desempeño Institucional de Información estadística contemplada en MiPG.  Diseñar e implementar un conjunto de procesos utilizando los datos producidos o administrados por las diferentes dependencias del Ministerio de Tecnologías de la Información y Comunicaciones.  La construcción de una herramienta que presente de forma interactiva distintos indicadores producidos con la información de los registros administrativos, realizando análisis de estos.  Información clara y de alta calidad, de los diferentes programas e iniciativas que desarrollan las áreas misionales del Ministerio, que permitan construir documentos de análisis y proyecciones de dichos programas o iniciativas, así como mostrar información sobre la gestión de la entidad. |

**Nota: Con la presentación de esta propuesta de proyecto nos comprometemos a dar respeto a todos y cada uno de los términos previstos en la Ley de protección de datos descritas en la Ley 1581 de 2012 así como las normas que la complementen, adicionen o sustituyan.**

1. <https://www.mintic.gov.co/portal/inicio/Ministerio/Acerca-del-MinTIC/>. Fecha de consulta 8 de mayo de 2020. [↑](#footnote-ref-2)
2. Conjunto de datos que contiene la información recogida y conservada por entidades u organizaciones en el cumplimiento de sus funciones o competencias misionales. DANE. Plan Estadístico Nacional. Bogotá año 2017. [↑](#footnote-ref-3)