|  |
| --- |
| **Formulario Evaluación Data Reto Integrativo** |

Muchas gracias por participar en el proceso de evaluación de organizaciones para el Reto Integrativo.

Por favor les pedimos hacernos llegar los siguientes documentos **hasta el martes 28 de setiembre al correo:** [**lbailetti@aporta.org.pe**](mailto:lbailetti@aporta.org.pe)

* **Formulario Evaluación de Data Reto Integrativo completo**
* **1 muestra de cada fuente de datos presentada en el formulario**

**(debe mostrar los campos principales y 5 registros mínimo, anonimizada, puede ser en CSV o Excel)**

* **Diccionario de datos, de no contar con uno les pedimos una breve descripción de variables que puedan ser difíciles de entender al revisar la tabla (p.ej. perfiles, categorías internas, otros)**
* **Brochures, links a información relevante sobre los proyectos activos de su organización**

**Si tienen alguna consulta, no duden en comunicarse con nosotros.**

1. **Datos de contacto**

|  |  |  |  |
| --- | --- | --- | --- |
| **Nombre de la organización** | Fundación Peruana de Cáncer | | |
| **Principales proyectos/servicios activos** | Asistencia integral al paciente especialmente mediante nuestro Albergue, educación en la prevención del cáncer e iniciativas de donaciones y becas educativas enfocadas en la descentralización de la atención oncológica. | | |
| **Nombre persona de contacto** |  | | |
| **Correo contacto** | Ingresar | **Celular contacto** | Ingresar |



**BREIT** |Programa Avanzado en Ciencia de Datos y Habilidades Globales

1. **Descripción de principales fuentes de datos**
2. **Describa brevemente las principales fuentes de datos de la organización/proyecto/servicio (p.ej: Datos demográficos de beneficiarios)**

**Fuente de datos 1**

|  |  |  |  |
| --- | --- | --- | --- |
| Nombre fuente de datos | HISMINSA (Health Information System del Ministerio de Salud) | | |
| Breve descripción | Sistema de registro de atenciones médicas en los centros de salud públicos | | |
| Fecha creación | 2019 | Última fecha actualización | 2023 |
| Frecuencia de actualización | Mensual | Esta frecuencia es regular | Sí |
| Breve descripción del procedimiento del recojo de la data (p.ej. si es por encuesta, dar información de ficha técnica, si es transaccional explicar) | Es un información transaccional. La unidad de análisis es la atención médica. | | |
| Número aprox. total de registros | 22 millones de atenciones por año (referencia 2021) | | |
| Tipo de datos (p.ej. fotos, videos, texto) | Datos estructurados en tablas | | |

**Fuente de datos 2**

|  |  |  |  |
| --- | --- | --- | --- |
| Nombre fuente de datos | SINADEF (Sistema Informático Nacional de Defunciones) | | |
| Breve descripción | Aplicativo informatico donde se registra los datos de las personas fallecidas en todo el Perú | | |
| Fecha creación | 2017 | Última fecha actualización | 2023 |
| Frecuencia de actualización | Diaria | Esta frecuencia es regular | Sí |
| Breve descripción del procedimiento del recojo de la data (p.ej. si es por encuesta, dar información de ficha técnica, si es transaccional explicar) | Es información de registros administrativos. La unidad de análisis es a nivel de persona. | | |
| Número aprox. total de registros | 248 mil personas por año (referencia 2021) | | |
| Tipo de datos (p.ej. fotos, videos, texto) | Datos estructurados en tablas | | |

1. **Describa brevemente las principales fuentes de datos de la organización/proyecto/servicio (p.ej: Datos demográficos de beneficiarios)**

**Fuente de datos 3**

|  |  |  |  |
| --- | --- | --- | --- |
| Nombre fuente de datos | ENDES (Encuesta Nacional Demográfica y de Salud) | | |
| Breve descripción | Información sobre la dinámica demográfica, el estado de salud en general, así como del estado y factores asociados a las enfermedades no transmisibles y transmisibles, y acceso a servicios de diagnóstico y tratamiento | | |
| Fecha creación | 1996 (utilizaríamos la data desde 2018) | Última fecha actualización | 2022 |
| Frecuencia de actualización | Anual | Esta frecuencia es regular | Sí |
| Breve descripción del procedimiento del recojo de la data (p.ej. si es por encuesta, dar información de ficha técnica, si es transaccional explicar) | Es información registrada por encuesta. | | |
| Número aprox. total de registros | 34 mil personas por año (referencia 2021) | | |
| Tipo de datos (p.ej. fotos, videos, texto) | Datos estructurados en tablas | | |

**Fuente de datos 4**

|  |  |  |  |
| --- | --- | --- | --- |
| Nombre fuente de datos | ENAHO (Encuesta Nacional de Hogares) | | |
| Breve descripción | Información orientada a dar seguimiento de los indicadores sobre las condiciones de vida en la población, incluyendo un módulo de salud | | |
| Fecha creación | 2004 | Última fecha actualización | 2017 |
| Frecuencia de actualización | Trimestral | Esta frecuencia es regular | Sí |
| Breve descripción del procedimiento del recojo de la data (p.ej. si es por encuesta, dar información de ficha técnica, si es transaccional explicar) | Es información registrada por encuesta. | | |
| Número aprox. total de registros | 121 mil personas por año (referencia 2021) | | |
| Tipo de datos (p.ej. fotos, videos, texto) | Datos estructurados en tablas | | |

1. **Describa brevemente las principales fuentes de datos de la organización/proyecto/servicio (p.ej: Datos demográficos de beneficiarios)**

**Fuente de datos 5**

|  |  |  |  |
| --- | --- | --- | --- |
| Nombre fuente de datos | CENARES (Centro Nacional de Abastecimiento de Recursos Estratégicos en Salud) | | |
| Breve descripción | Investigación estadística, una de las mas importantes, que se ejecuta de manera continua en el INEI, contiene información acerca de la salud materna e infantil, prevalencia anticonceptiva, fecundidad y mortalidad de la población, conocimiento del VIH-SIDA, estatus de la mujer y violencia doméstica. Asimismo, contiene indicadores de resultado final e intermedio de tres programas estratégicos: Articulado Nutricional, Salud Materno Neonatal; y, Acceso de la Población a la Identidad. | | |
| Fecha creación | 2011 | Última fecha actualización | 2023 |
| Frecuencia de actualización | Mensual | Esta frecuencia es regular | Sí |
| Breve descripción del procedimiento del recojo de la data (p.ej. si es por encuesta, dar información de ficha técnica, si es transaccional explicar) | Es información de registros administrativos. | | |
| Número aprox. total de registros | 136 registros por año con información agregada por dependencia, punto de destino, mes, lote, fecha de vencimiento (referencia 2021) | | |
| Tipo de datos (p.ej. fotos, videos, texto) | Datos estructurados en tablas | | |

**Fuente de datos 6**

|  |  |  |  |
| --- | --- | --- | --- |
| Nombre fuente de datos | Ingresar | | |
| Breve descripción | Ingresar | | |
| Fecha creación | Desde qué año existe esta fuente de datos | Última fecha actualización | Desde qué año existe esta fuente de datos |
| Frecuencia de actualización | Ingresar | Esta frecuencia es regular | Ingresar |
| Breve descripción del procedimiento del recojo de la data (p.ej. si es por encuesta, dar información de ficha técnica, si es transaccional explicar) | Ingresar | | |
| Número aprox. total de registros | Ingresar | | |
| Tipo de datos (p.ej. fotos, videos, texto) | Ingresar | | |

**🡪 Si cuentan con más fuentes de datos, pueden llenar otro formulario incluyendo las demás fuentes.**

1. **¿Las fuentes de datos se encuentran relacionadas? ¿Cómo?**

Todas las fuentes tienen como dato relacional el ubigeo (código de ubicación geográfica para identificar la región-provincia-distrito) por cada registro realizado.

1. **¿Además de estas fuentes de datos, tienen otro tipo de información no estructurada almacenada digitalmente que sea relevante para sus proyectos/servicios?** (p.ej. transcripciones, grabaciones, e-mails, fotografías)

No

1. **¿Existe alguna fuente de datos externa que sea relevante para su proyecto que se pueda disponer para el proyecto?** (p.ej. data proporcionada por el MINEDU o alguna organización externa a ustedes)

Las fuentes de datos recolectadas para el presente proyecto son fuentes de datos externas y públicas. Se detalla el origen de cada fuente de datos en las siguientes líneas:

* MINSA (Ministerio de Salud): Llevan a cabo el registro del HISMINSA (Health Information System del Ministerio de Salud) y el SINADEF (Sistema Informatico Nacional de Defunciones).
* INEI (Instituto Nacional de Estadística e Informática): Lleva a cabo el Censo 2017, ENAHO (Encuesta Naciona de Hogares), ENDES (Encuesta Nacional Demográfica y de Salud Familiar).
* CENARES (Centro Nacional de Abastecimiento de Recursos Estratégicos en Salud): Lleva a cabo el registro interno que contiene el mismo nombre.

1. **¿Hay algún proyecto en los próximos 12 meses que vaya a cambiar de forma importante la estructura, la forma de recojo o características de los datos? Por favor explique el proyecto y el impacto que tendría en los datos**

No, no hay proyecto alguno.

1. **¿Alguna especificación adicional que considere relevante sobre las fuentes declaradas?**

Ninguna

1. **Sobre las posibles soluciones a desarrollar en el Challenge Integrativo**

**De ingresar al Challenge Integrativo, el equipo de nuestros alumnos, profesionales relacionados a analítica de datos diseñaran en conjunto con los representantes de su organización la solución a desarrollar. Sin embargo, de manera preliminar quisiéramos saber:**

**¿Qué necesidades/dolores actuales de su organización cree que podría ayudar nuestro equipo de alumnos y expertos a aliviar durante el Challenge Integrativo? ¿Tienen algunas soluciones potenciales en mente que creen se podrían trabajar (por favor listarlos)?** (esta es una pregunta de levantamiento preliminar, se pueden plantear ideas o necesidades generales, no es requisito tener una solución definida para ingresar al Challenge)

- Toda la data que almacena el Observatorio del cáncer de la FPC proviene de fuentes públicas; sin embargo, es importante resaltar que en algunos casos existe un subreporte. En un caso particular, con la fuente del SINADEF, se identificó que un periodo largo deL reporte no es confiable (09/22 a 08/23) debido a que en setiembre 2022 se notificó la existencia de actas de defunciones irregulares lo cual ocasionó que el sistema electrónico de defunciones se suspendiera. Por ese motivo, lo que buscamos resolver con el apoyo del equipo asignado al Challenge Integrativo de APORTA es solucionar el problema del subreporte de esta fuente específica del SINADEF con el desarrollo de una solución o programa estadístico que logre estimar proyecciones de la mortalidad por cáncer. Una solución que ya provee por ejemplo la IARC (International Agency for Research on Cancer) – Globocan.

- Cuanto más temprano es detectado el cáncer, las posibilidades de vencerlo alcanzan hasta el 90%; sin embargo, 2/3 se detectan en estadios avanzados por lo cual uno de nuestros pilares es la educación en la prevención y detección temprana del cáncer. Existen muchas variables que pueden condicionar que una persona desarolle cáncer. El análisis de la data del Observatorio de la FPC también puede ayudarnos potencialmente a establecer correlaciones estadísticas e incluso de causalidad entre diferentes variables (factores de riesgo, comorbilidades, nivel de pobreza, accesos sistema sanitario, entre otros) que nos ayuden a predecir la posibilidad de diversos tipos de cáncer en comunidades específicas.

1. **Evaluación de la organización**

**Marque el nivel de desarrollo en el que se encuentra su organización/proyecto en cada dimensión**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Dimensión** | **Nivel** | | | |
| **Básico** | **Intermedio** | **Avanzado** | **Líder** |
| **Cultura de datos** | El staff sabe qué datos hay  pero no reconocen su  importancia. | Pocos individuos saben qué  data hay y qué se puede  hacer con ella. | La organización tiene una idea clara de cómo los datos  pueden servir para tomar  decisiones de negocio | La organización tiene una  cultura de datos y demanda  que las decisiones se  justifiquen con datos |
| **Recolección**  **de datos** | El staff provee datos  esporádicamente porque  están obligados a hacerlo. | El staff regularmente  provee datos porque están  obligados a hacerlo. | El staff regularmente  provee datos y  genera insights | El staff provee datos en  tiempo real y decide en  función a los insights |
| **Importancia para**  **el liderazgo** | Los líderes no entienden  cómo los datos pueden  ayudar a cumplir la misión de  la organización. | Los líderes quieren utilizar  los datos pero no saben  cómo exactamente | Los líderes tienen una idea  clara de cómo los datos  pueden dirigir las  decisiones de negocio. | Los líderes crean una cultura de datos en la organización y  demandan que las decisiones se justifiquen con datos |
| **Recursos humanos** | El staff mantiene sus datos  en silos. | El staff puede ver el valor  de los datos pero no tiene  la capacidad para crear  insights valiosos. | El staff ve bien el valor de los  datos, pero no tiene las  habilidades suficientes o  infrastructura para  crear insights valiosos | La organización tiene un staff  dedicado a la gestión de la  data, almacenamiento, rigor y  limpieza. |
| **Políticas de uso**  **de datos** | No existen políticas para el  uso, transferencia y  compartir de la data | Hay políticas para el uso,  transferencia y compartir de  la data pero no cubre toda  la data que existe | Hay buenas políticas para el  uso, transferencia y  compartir de la data  internamente. | Hay buenas políticas para el  uso, transferencia y  compartir de la data interna  y externamente |
| **Probabilidad de**  **aplicación de soluciones usando data analytics** | Es poco probable que comencemos a usar data analytics a partir de esta experiencia | Es probable que incorporemos análisis básicos de data analytics a partir de esta experiencia | Es probable que esta experiencia nos ayude a alcanzar formas más avanzadas de gestionar la data | Esta experiencia aportará una solución más de las varias soluciones que tenemos utilizando data science |
| **Dimensión** | **Nivel** | | | |
| **Básico** | **Intermedio** | **Avanzado** | **Líder** |
| **Accesibilidad** | La data solo es accesible desde la fuente que la genera | La data es accesible desde fuera pero en un formato que nos genera algunos problemas | La data es accesible desde fuera con un formato estándar | La data es accesible desde un API en un formato estándar |
| **Almacenamiento** | Papel | PDF, Excel o bases de datos con formato propietario | Bases de datos y CSV, JSON, TXT o archivos de texto plano | Bases de datos que permiten llamadas desde un API |
| **Integración** | La data sólo se encuentra donde se genera | La data se exporta ocasionalmente | La data se agrega en un data warehouse | La data se integra con información externa a un data warehouse |
| **Relevancia y suficiencia** | La data es irrelevante al objetivo de proyecto | La data es relevante pero es inconsistente | La data es relevante pero no es suficiente | La data es relevante y suficiente |
| **Calidad** | La data tiene muchos valores vacíos, filas vacías y errores | La data tiene filas vacías y posibles errores | La data tiene pocos valores vacíos y pocos errores | La data no tiene valores vacíos, baja posibilidad de error y procesos de validación |
| **Datos temporales** | Solo se mantiene la data actual, no se acumula data histórica | La data histórica es almacenada, pero se sobre-escriben algunos valores | La data histórica es almacenada y la nueva se agrega a la antigua | La data histórica es almacenada y los schemas se mapean de manera consistente |
| **Documentación** | No hay documentación de las variables, pero se pueden crear diccionarios | Ya existen diccionarios | Los diccionarios tienen información sobre las condiciones bajo las que se capturó la data | Los diccionarios tienen información sobre los supuestos y posibles sesgos de la data |