



Date: 7/26/2023

To: **All Bidders**

From: **International Development Law Organization, Headquarters Procurement Unit**

Subject:

Clarification No: 2

Reference is made to IDLO RFP HN_2023-000035 issued on July 13, 2023 for the procurement of Software for data anonymization through IDLO Tenders website.

Please find hereunder the queries received and IDLO's response to them.

Preguntas técnicas:

1. Entendemos que el aplicativo a desarrollar debe ser web. ¿Se puede proponer un framework de front end como devextreme que se integre a .NET? Este framework permite mejorar la experiencia de usuario con componentes como reportes, tablas con ordenamiento y búsqueda, dashboards, entre otros.?
R. El aplicativo a desarrollar no es web, solo opera en modo local a partir del archivo que ingresa el operador y descargando la versión anonimizada en la misma PC en que se está operando. Los archivos auxiliares deben estar en red interna, o sea solo accesibles para todas las versiones del aplicativo a desarrollar instaladas en las PC de la red local. En caso de utilizar componentes de tercero estos deben ser libres, DEVEXTREME requiere de licenciamiento y no contar con ello imposibilitaría el mantenimiento futuro de la aplicación.
Todos los componentes deben ser de libre uso o bien desarrollados específicamente para el software.
2. ¿Qué base de datos se debe utilizar? ¿Podemos proponer un motor de bases de datos (SQL Server, Oracle, MySQL)? ¿Se cuenta con la licencia respectiva del motor seleccionado?
R. Solo los archivos auxiliares podrían considerarse para organizarlos como base de datos; incluso podrían ser archivos planos. Estos archivos son solo utilizados por las múltiples versiones del aplicativo a desarrollar, pero si deberían poder ser editables fácilmente (como eliminar entradas o corregir el texto de alguna entrada).
3. ¿Dónde se hospedará la herramienta implementada? ¿En servidores propios o en una nube como Amazon Web Services? Este es un dato importante dado que si se utilizará una nube se pueden aprobar elementos de infraestructura existentes.
R. El software debe poder ser instalada de forma individual en el disco C de computadoras y poder ser utilizado de forma autónoma. El software no debe ser desarrollado para ser utilizado en línea o a través de una nube, ni subir datos a ninguna nube.

4. Como se prevé manejar la autenticación. ¿Integrada a un LDAP como Active Directory, Cognito de AWS o con un catálogo propio de usuarios de la aplicación?
- R. A través de un catálogo propio.
5. ¿Qué herramienta de clean code se debe proponer?
- R. Se espera que el consultor al momento de codificar aplique técnicas de clean code, por ejemplo en el nombrado de variables, funciones, métodos, código documentado, etc. Lo que se busca es tener un código legible e intuitivo para el mantenimiento.
6. Se cuenta con algún requerimiento de arquitectura que se deba conocer. Por ejemplo, microservicios, Modelo Vista Controlador, etc.
- R. Microservicio

Preguntas funcionales:

7. Al leer los Términos de Referencia entendemos que la herramienta transformaría archivo por archivo. Es decir, cada proceso de transformación recibe como entrada un archivo para generar una salida. ¿Se requiere de un proceso masivo?
- R. No se prevé análisis masivos de archivos estructurados de texto. El software deberá analizar archivo por archivo de forma automática y de forma manual por el operador o usuario.
8. En los Términos de Referencia se menciona que el operador es el encargado ingresar el archivo a transformar. ¿Entendemos que el operador es un rol de usuario que usa la herramienta?
- R. Efectivamente, el usuario u operador será quien ingresará el archivo al software para su análisis y deberá ser capaz de tomar decisiones sobre el proceso de anonimización realizado por el software. Finalmente es el operador quien descarga la versión anonimizada.
9. En los términos de referencia solo se menciona al operador. ¿Hay otros roles involucrados?
- R. Se plantea que el software sea utilizado por múltiples usuarios (operadores) y a manera de mantenimiento por un supervisor de la Dirección de Infotecnología quien solo podrá editar los archivos auxiliares y diccionarios.
10. Además de nombre y apellido, telephone numbers, DNI (National Identification Document. que otro dato es sensitivo y requiera transformación?
- R. Cualquier otro dato que pueda llevar a identificar a una persona. Tal como apodosos o sobre nombres, lugar de procedencia, lugar de trabajo, lugar de estudios. En esta operación de selección de datos a anonimizar, el usuario u operador tendrá la potestad de elegir el dato a anonimizar.
11. ¿Que son los archivos auxiliares y catálogos permanentes que mencionan en la pág. 17 del RFP?
- R. Existen varias estrategias para lograr que el software solicitado logre los resultados esperados: según sea la estrategia es posible que se utilicen archivos auxiliares o diccionarios. Algunos ejemplos de posibles archivos auxiliares son:
1. "Palabras de arranque de frase". Algunos software de anonimización suelen utilizar un archivo auxiliar o diccionario de "palabras de arranque"; por ejemplo "Que", "No", "El", etc. junto con una estrategia para detectar comienzos de frase (por ejemplo palabras en mayúscula precedidas por los caracteres: <punto

.><espacio>, <CR>, <numero>< paréntesis >, etc.). Este archivo o diccionario se irá creando y enriqueciendo paulatinamente durante el desarrollo y los usos sucesivos. Si la propuesta incluye esta posibilidad se espera que la versión primigenia contenga suficientes entradas como para funcionar sin complicaciones. Podría ser conveniente (para el supervisor) conocer frecuencia de aparición de estas palabras (durante el uso) como información útil para ser editadas.

2. Una base de datos con los nombres, apellidos y fecha de nacimiento de los nacionales y residentes. Si se selecciona esta estrategia la base de datos debe ser provista junto con el software, más un mecanismo de actualización y sus costos.

12. Para que se sirven los “Positive file/catalog”? Entendemos que son universales, es decir de referencia para cualquier proceso de transformación.

R. Este archivo/diccionario incluye palabras que al encontrarlas indican que el archivo que se ingresa tiene una probabilidad significativa de contener datos confidenciales que impliquen o requieran que algunas personas vinculadas deban anonimizarse (por ejemplo este archivo incluirá palabras como: sexual, incesto, violación, y una lista de enfermedades o síntomas). El software debe incluir la posibilidad de que el operador ingrese una palabra o grupos de palabras a este archivo (en este caso el supervisor deberá recibir alertas periódicas sobre las palabras ingresadas y este archivo deberá mantener marcadas las palabras aun no validadas por el supervisor). Este archivo podría entregarse con las palabras que se utilizaron en la fase de desarrollo (sujeto a revisión del supervisor). Estos archivos son de referencia para todos los procesos de anonimización que realizan todos los operadores (o sea son acumulativos).

13. ¿Para qué sirven los negative files?

R. Los archivos/diccionarios negativos son una recolección de palabras que no se identificarán ni detectarán (y por ende no se anonimizarán) aun estando escritas en mayúsculas (si es que la estrategia del software se basa en la búsqueda de palabras que inician con mayúsculas). Por ejemplo: Constitución, Honduras, Pacto, Internacional, Derechos, Civiles, Políticos, etc. Este diccionario incluye también los nombres y apellidos de los jueces, los fiscales, defensores públicos, autoridades nacionales, etc. Puede contener palabras simples o compuestas (e.g. no sería conveniente que “Martínez” esté en este diccionario, pero si “Emhed Salomé Martínez Padilla” que fue Secretario de Transparencia).

Este, como los restantes diccionarios se comienzan a generar en las pruebas y luego se van completando e incrementando con el uso.

14. ¿Para qué sirven los archivos estructurados? ¿Podrían brindar un ejemplo de este tipo de archivos?

R. Cuando hablamos de los archivos estructurados nos referimos a la información que se suele encontrar en la mayoría de las bases de datos, en este caso nos referimos a archivos en formatos PDF, Word o texto. Serán archivos de tipo texto que se suelen mostrar en filas y párrafos con títulos y subtítulos. Son archivos que pueden ser ordenados y procesados fácilmente por todas las herramientas de procesamiento de datos, en especial bajo ambiente Windows.

15. ¿Qué servicios debe incluir la API de integración con SEJE?

R. Vía una API se debe enviar un número de expediente (el formato esta estandarizado por la Dirección de Infotecnología), como respuesta se deben recibir los siguientes datos de todas las personas registradas como intervinientes en ese expediente: Nombres; Apellidos, tipo de parte (condición como parte en el proceso), fecha de nacimiento. La propuesta debe definir los requisitos para reconocer las respuestas

(por ejemplo en un archivo JSON). Normalmente Nombres es todos los nombres separados por espacio, igual ocurre con Apellidos, tipo de parte (toma valores como actor, demandado, fiscal, víctima, ofendido, niño motivo de protección, querellante, abogado defensor privado, abogado defensor público, testigo, testigo protegido, perito forense, perito de parte, tercero, etc.) Todos estos datos (cuando están accesibles) deben dar más certeza al proceso de anonimización y advertir al operados los que son datos confidenciales (por ejemplo víctimas o niños).

16. Comprendemos que se analizarán unstructured files (doc, .docx, .txt, .rtf,). ¿No nos queda claro cuál es la data estructurada que debe analizarse? ¿Podrían proveer un ejemplo?
- R. Los archivos a analizarse serán archivos de texto con filas, párrafos con títulos y subtítulos. No contienen normalmente tablas ni imágenes incrustadas. Un ejemplo es una sentencia condenatoria de un delito penal. Véase Adjunto. Rara vez contienen recursos más allá de “negritas” o “subrayados”. Una posibilidad observada (pero remota) es que alguna vez un archivo incluya tablas; sería un plus si el software las interpretara como si fueran texto plano —pero esto no es un requisito. Para cualquier otra rareza (imágenes, por ejemplo, o sonido) el operador deberá suprimirlas antes del ingresarlas al software y luego proceder manualmente para insertarlas en la versión final anonimizada.
17. ¿En el Paso de Configure, para qué sirven los Privacy model y el coding model? ¿Pueden brindarnos un ejemplo?
- R. Se refiere a las opciones de reemplazo de nombres y apellidos por iniciales o niño2, o victima1 según sea el caso. La cantidad de iniciales (2, 3 o 4) como el uso de parte1 depende de los riesgos de re-identificación y deben ser decididas por el operador;
18. ¿De qué tamaño son los archivos para analizar?
- R. Los archivos a analizar serán archivos de texto (normalmente textos de sentencias judiciales, o sea no contienen normalmente tablas ni imágenes incrustadas) los tamaños son variables, difícilmente superen 1MB. Como ejemplo del tipo del archivo que se ingresa puede tomarse la segunda parte del archivo (en formatos (doc, .docx, .txt, .rtf,). — no en .pdf) con se descarga aquí:
<https://jurisprudencia.poderjudicial.gob.hn/ConsultaHN/PDFAmparo.aspx?opcion=1®=15071>
Desde donde dice: SENTENCIA – CERTIFICACION
19. En la página 16 se menciona “e.g., focusing on machine learning tasks”. ¿Podrían aclarar que se espera sobre machine learning?
- R. De acuerdo a las especificaciones técnicas la propuesta del software que se espera debe realizar la detección de todas situaciones en las que el operador debe decidir (el software debe minimizar errores de detección), luego en la pantalla de análisis/decisión el software debe maximizar información y alternativas necesarias en el proceso de decisión (opciones de anonimización). Sobre la pregunta formulada, obviamente que la propuesta incluya una perspectiva de análisis estadístico, o de otro tipo, sobre la forma en que los operadores toman sus decisiones y que este análisis pueda incorporarse a un procedimiento automatizado (de información o decisión) sería considerado como muy positivo, pero no es un requisito esperado que se incluya inteligencia artificial.

Preguntas de logística:

20. ¿Se tienen definidas cuales son las reuniones presenciales? ¿O podemos definirlo en la propuesta?
- R. Efectivamente, las reuniones están sujetas a propuesta de cualquiera de las partes,

- pudiendo en la propuesta inicial colocar reuniones de acuerdo a las necesidades de ejecución.
21. ¿La reunión semanal es presencial o se puede proponer virtual? Recomendamos maximizar las reuniones virtuales por el tiempo y costo que implican los traslados y el ajustado tiempo del proyecto.
- R. La frecuencia de las reuniones se definirá en el transcurso de la ejecución de la consultoría y podrán realizarse tanto de manera virtual como presencial.
22. ¿El plazo de 60 días se trata de días naturales o hábiles? ¿Se puede sugerir un plan de trabajo mayor a 60 días? Consideramos ajustado el plazo propuesto.
- R. El plazo son 60 días naturales contados a partir de la firma del contrato.

For your consideration while preparing your response to the above mentioned RFP. All other terms and conditions of the tender remain unchanged and remain applicable in full force.

We appreciate your continuous interest with IDLO and looking forward to receiving your valued offer.

Headquarters
Procurement Unit
IDLO

Headquarters
Viale Vaticano, 106
00165 Rome, Italy
Tel + 39 06 4040 3200
Fax + 39 06 4040 3232
idlo@idlo.int

Branch Office
Hofweg 9E
2511AA The Hague
The Netherlands
Tel + 31 70 240 0870
thehague@idlo.int

Office of the Permanent Observer
to the United Nations
336 East 45th Street, 11th floor
New York, NY, 10017, USA
Tel + 1 212 867 9707
newyork@idlo.int

Office of the Permanent Observer
to the United Nations
23 av. de France
1202 Geneva, Switzerland
Tel + 41 22 734 41 40
geneva@idlo.int

idlo@idlo.int
www.idlo.int
 @IDLO
 IDLOnews