Date: 7/26/2023

To: **All Bidders**
From: **International Development Law Organization, Headquarters Procurement Unit**
Subject:

<u>**Clarification No: 2**</u>

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.

Technical questions:

1. We understand that the application to be developed must be web. Can you propose a front end framework like devextreme that integrates with .NET? This framework allows you to improve the user experience with components such as reports, tables with ordering and search, dashboards, among others?
A. The application to be developed is not a web application, it only operates locally from the file entered by the operator and downloading the anonymized version on the same PC on which it is operating. The auxiliary files must be on the internal network, that is, only accessible to all versions of the application to be developed installed on the PCs on the local network.
    In the case of using third-party components, these must be free, DEVEXTREME requires licensing and not having it would make future maintenance of the application impossible.
    All components must be free to use or developed specifically for the software.

2. What database should be used? Can we propose a database engine (SQL Server, Oracle, MySQL)? Do you have the respective license for the selected engine?
A. Only the auxiliary files could be considered to organize them as a database; they could even be flat files. These files are only used by the multiple versions of the application to be developed, but they should be able to be easily edited (such as deleting entries or correcting the text of any entry).

3. Where will the deployed tool be hosted? On your own servers or in a cloud like Amazon Web Services? This is an important piece of information given that if a cloud is to be used, existing infrastructure elements can be approved.
A. The software must be able to be installed individually on the C drive of computers and be able to be used autonomously. The software must not be developed to be used online or through a cloud, nor upload data to any cloud.

4. How authentication is expected to be handled. Integrated with an LDAP such as Active Directory, AWS Cognito or with your own catalogue of application users?
A. Through our own catalogue.

5. What clean code tool should be proposed?
A. The consultant is expected to apply clean code techniques when coding, for example in the naming of variables, functions, methods, documented code, etc. What is sought is to have a readable and intuitive code for maintenance.

6. There are some architectural requirements that should be known. For example, microservices, Model View Controller, etc.
A. Microservice

Functional questions:

7. By reading the Terms of Reference we understand that the tool would transform file by file. That is, each transformation process receives a file as input to generate an output. Is a massive process required?
A. Mass analysis of structured text files is not foreseen. The software shall analyze file by file automatically and manually by the operator or user.

8. In the Terms of Reference it is mentioned that the operator is in charge of entering the file to be transformed. Do we understand that the operator is a user role that uses the tool?
A. Indeed, the user or operator will be the one who will enter the file into the software for analysis and must be able to make decisions about the anonymization process carried out by the software. Finally it is the operator who downloads the anonymized version.

9. In the terms of reference only the operator is mentioned. Are there other roles involved?
A. It is proposed that the software be used by multiple users (operators) and as maintenance by a supervisor from the Infotechnology Department who will only be able to edit the auxiliary files and dictionaries.

10. In addition to name and surname, telephone numbers, DNI (National Identification Document. What other data is sensitive and requires transformation?
A. Any other data that can lead to identify a person. Such as nicknames or names, place of origin, place of work, place of study. In this data selection operation to be anonymized, the user or operator will have the power to choose the data to be anonymized.

11. What are the auxiliary files and permanent catalogs mentioned on p. 17 of the RFP?
A. There are several strategies to ensure that the requested software achieves the expected results: depending on the strategy, auxiliary files or dictionaries may be used. Some examples of possible auxiliary files are:
1. "Phrase starter words." Some anonymization software often uses an auxiliary file or dictionary of "startup words"; for example "Que", "No", "El", etc. along with a strategy for detecting sentence beginnings (for example, uppercase words preceded by the characters: <period .><space>, <CR>, <number>< parentheses)>, etc.). This file or dictionary will gradually be created and enriched during development and successive uses. If the proposal includes this possibility, it is expected that the original version

contains enough entries to work without complications. It could be convenient (for the supervisor) to know the frequency of appearance of these words (during use) as useful information to be edited.

2. A database with the names, surnames and date of birth of nationals and residents. If this strategy is selected, the database must be provided together with the software, plus an update mechanism and its costs.Para que se sirven los "Positive file/catalog"?

12. We understand that they are universal, that is, of reference for any transformation process.

A. This file/dictionary includes words which, when found, indicate that the file being entered has a significant probability of containing sensitive data that implies or requires that some associated individuals must be anonymized (for example, this file will include words such as: sexual, incest, rape, and a list of diseases or symptoms). The software must include the possibility for the operator to enter a word or groups of words to this file (in this case the supervisor must receive periodic alerts about the words entered and this file must keep marked the words not yet validated by the supervisor). This file could be delivered with the words that were used in the development phase (subject to supervisor review). These files are of reference for all the anonymization processes carried out by all the operators (in other words, they are cumulative).

13. What are the negative files for?

A. Negative files/dictionaries are a collection of words that will not be identified or detected (and therefore will not be anonymized) even if they are written in uppercase (if the software's strategy is based on searching for words starting with uppercase). For example: Constitution, Honduras, Covenant, International, Rights, Civil, Political, etc. This dictionary also includes the names and surnames of judges, prosecutors, public defenders, national authorities, etc. It may contain simple or compound words (e.g. it would not be convenient for "Martínez" to be in this dictionary, but "Emhed Salomé Martínez Padilla" who was Secretary of Transparency if).
This, like the other dictionaries, begins to be generated in the tests and then they are completed and increased with use.

14. What are structured files for? Could you provide an example of this type of files?

A. When we talk about structured files, we are referring to the information that is usually found in most databases, in this case we are referring to files in PDF, Word or text formats. They will be text type files that are usually displayed in rows and paragraphs with titles and subtitles. They are files that can be easily ordered and processed by all data processing tools, especially under Windows environment.

15. What services should the SEJE integration API include?

A. Via an API, a file number must be sent (the format is standardized by the Infotechnology Department), in response the following data must be received from all the people registered as participants in that file: Names; Surname, type of party (condition as a party in the process), date of birth. The proposal must define the requirements to recognize the responses (for example in a JSON file). Normally Names are all names separated by spaces, the same happens with Surnames, type of party (takes values such as plaintiff, defendant, prosecutor, victim, offended, child protected, plaintiff, private defense attorney, public defense attorney, witness, protected witness, forensic expert, party expert, third party, etc.) All these data (when accessible) should give more certainty to the anonymization process and warn the operator that the data is confidential (for example, victims or children).

16. We understand that unstructured files (doc, .docx, .txt, .rtf,) will be analyzed. Is it not clear to us what is the structured data that must be analyzed? Could you provide an example?

A. The files to be analyzed will be text files with rows, paragraphs with headings and subheadings. They do not normally contain tables or embedded images. An example is a conviction for a criminal offense. See Attached. They rarely contain resources beyond "bold" or "underlined". An observed (but remote) possibility is that a file ever includes tables; it would be a plus if the software interpreted them as if they were plain text — but this is not a requirement. For any other rarity (images, for example, or sound) the operator must delete them before entering them into the software and then proceed manually to insert them into the final anonymized version.

17. In the Configure Step, what are the Privacy models and the coding models for? Can you give us an example?

A. It refers to the options to replace names and surnames with initials or child2, or victim1 as the case may be. The number of initials (2, 3 or 4) as the use of part1 depends on the risks of re-identification and must be decided by the operator.

18. How big are the files to analyze?

A. The files to be analyzed will be text files (usually texts of judicial sentences, that is, they do not normally contain tables or embedded images) the sizes are variable, hardly exceeding 1MB. As an example of the type of the file that is entered, the second part of the file can be taken (in formats (doc, .docx, .txt, .rtf,). — not in .pdf) with download here:
https://jurisprudencia.poderjudicial.gob.hn/ConsultaHN/PDFAmparo.aspx?opcion=1&reg=15071
From where it says: JUDGMENT – CERTIFICATION

19. On page 16 there is a mention of "e.g., focusing on machine learning tasks". Could you clarify what is expected about machine learning?

A. According to the technical specifications, the software proposal that is expected to detect all situations in which the operator must decide (the software must minimize detection errors), then on the analysis/decision screen, the software must maximize information and alternatives necessary in the decision process (anonymization options). Regarding the question asked, obviously the proposal includes a statistical or other analysis perspective on the way in which operators make their decisions and that this analysis can be incorporated into an automated procedure (of information or decision) would be considered very positive, but it is not an expected requirement that artificial intelligence be included.

Logistics questions:

20. Have you defined which are the face-to-face meetings? Or can we define it in the proposal?

A. Indeed, the meetings are subject to a proposal by either party, and the initial proposal may place meetings according to the execution needs.

21. Is the weekly meeting face-to-face or can it be proposed virtual? We recommend maximizing virtual meetings due to the time and cost involved in travel and the tight project schedule.

A. The frequency of the meetings will be defined during the execution of the consultancy

and can be held both virtually and in person.

22. Is the 60-day period calendar days or business days? Can you suggest a work plan longer than 60 days? We consider the proposed deadline adjusted.
A. The term is 60 calendar days from the signing of the contract.

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