

Request for Information: Data Cleansing of the National Data Warehouse datasets using AI.

Date: 25th July 2024

Q1	RFI Form Submission: Can you confirm that the RFI form provided at https://eforms.modee.gov.jo/ is the correct platform for submitting our information and that we will receive approval from MODEE once our form is submitted?
A1	Yes, we confirm that the RFI form provided at https://eforms.modee.gov.jo/ is the correct platform for submitting your information. Upon completing and submitting the form, the last page will confirm your submission
Q2	Tender Document Access: We understand that we will be able to purchase the tender documents once we are qualified. Could you please confirm if this is the case or if the documents will be sent to us directly upon qualification?
A2	After the RFI process, the qualified bidder will receive the RFP directly free of charge.
Q3	Inquiries Related to the RFP: If we qualify and purchase the RFP, will we be able to ask any questions related to the RFP?
A3	All qualified bidders will be able to ask any questions related to the RFP document within a dedicated period.
Q4	Modifications to the Google Form: We have multiple options regarding the Academia AI researcher. Is it possible to make any changes to the Google form after submission if needed? Alternatively, can we apply again or submit our application through the attached document?
A4	Bidders must apply using the RFI form provided (https://eforms.modee.gov.jo/) and they will not be able to submit another form. Other way for submissions will not be accepted.
Q5	Use case 1 - Determine National ID: Is the data in English or Arabic or a mix of both?
A5	The Data is in Arabic Language
Q6	Use case 1 - Determine National ID: Is there any specific SQL technology (SQL Server, MySQL, Oracle...) being used for the data. Should the same technology be used for the Non-AI statements / scripts?
A6	No, there is no specific SQL technology used. You are not required to use the same technology for the non-AI statements/scripts.

Q7	Use case 1 - Determine National ID: You mention that you already experimented a matching with a sample of 330K records with a 82% accuracy. Why did you not extend the same script / logic to the rest of the 1.8 M records? Do you foresee any issues with the script used?
A7	Because we just worked with a sample Tawjihi dataset for the 2002-2004 batch. We plan to extend this work using the AI method.
Q8	Use case 1 - Determine National ID: Would you be sharing the script that was used in the experiment so that can check what was tried and extend it to rest of the data (out of sample / experiment data)
A8	No, we will not be able to share the script that we used in our experiment.
Q9	Use case 1 - Determine National ID: Are there other sources of data (other than CSPD) that contain National ID and which can be leveraged for the POC? If so, would the Department of Examinations and Tests be responsible to provide this to us?
A9	The primary source for the national ID number (single source of truth) is the CSPD, and we have to rely on it. In case there would be other reliable source of data <u>in addition to CSPD</u> , all stakeholders must discuss and agree on it including who will be responsible for the data provision.
Q10	Use case 1 - Determine National ID: In point "(6) General conditions of Use case 1", you have mentioned that we can exclude data with no assigned National ID. Are there columns in the table that will help us with the exclusion? Do you have an estimate of how many records might excluded in the 1.8 M records?
A10	No estimation at this time. The data will be provided for the winning bidder.
Q11	Use case 2: Linking municipality property data to DLS data Is the data in English or Arabic or a mix of both?
A11	The Data is in Arabic Language
Q12	Use case 2: Linking municipality property data to DLS data In point "(6) General conditions of Use case 2", you mention that we can use other RDBMS. Will you share a copy of the data from MOLA or DLS with us so that we can load it into the new RDBMS?
A12	The data will be provided for the winning bidder.
Q13	Use case 3: Anonymizing Court decisions: Are all the documents to be anonymized in Arabic?

A13	The Data is in Arabic Language
Q14	Use case 3: Anonymizing Court decisions: Are there any specific Arabic LLM you propose we use?
A14	No
Q15	Use case 3: Anonymizing Court decisions: Do you have GPUs if we have to re-train the model locally?
A15	No
Q16	In the Join POC to be managed by the consortium, who would be the leading party to execute the project. Who will be the Unified Point of Contact.
A16	MoDEE will be the leading party to execute the project As mentioned in the ToR document (Chapter-4), The Joint Team (Private IT company & Academic AI researcher) must have a single, unified point of contact (focal point) to report to and communicate with stakeholders (MoDEE, AI Steering Committee, JICA consultant team, target government organization) which represents both private IT company and academic AI researcher. Mostly the focal point will be a member of an IT private company.
Q17	Please confirm that identifying the Academia personnel (AI Researcher) and expertise is part of the scope of the technology vendor or is it a separate requirement.
A17	Identifying the Academia personnel (AI researcher) and expertise is a part of the scope of the technology vendor. During the RFI process, both the vendor and proposed Academia personal will be evaluated accordingly.