**Expression of Interest**

# **Project Details:**

|  |  |
| --- | --- |
| **Project Name** | DDaT25039 - Auto-Classification Tool |
| **Response required by** | 16:00 Friday 14th February 2025 |
| **Response required to** | DDaTProcurement@uksbs.co.uk |

# **Description of the Project**

|  |
| --- |
| The services required is the ability to classify the UKRI portfolio of applications and awards (Research Councils, Research England, Innovate UK) to a pre-existing classification system. The need for automating the classification is to reduce time spend on manual categorisation and improve operational efficiency. The solution is to classify new applications and our historic portfolio.  The classification systems required are: Sustainable Development Goals (SDG); OECD sector codes; Health Research Classification System (HRCS); Research, Condition and Disease Categorization (RC); Fields of Research (FoR); Common Scientific Outline and Cancer Type Codes; Units of Assessment (UoA).  In addition, we require a solution to identify Trusted Research and Innovation for a quarterly report to DSIT.  This is required to ensure continuity of delivering business critical reports to DSIT, IATI, Annual reporting to boards, and being able to answer requests of ‘what have we funded in defined research areas’, that have a quick turnaround for example PQs, FOIs.  **Aims & Objectives**  The aim of the tender exercise is to focus on acquiring an AI/ML solution that can automatically classify data into pre-existing structured classification systems, improving operational efficiency and reducing manual intervention. The solution must also allow access to other funder data, not just UKRI, enabling comparison, prioritisation and gap analysis. It must apply the same model to all this data, so we can compare like for like across funders, rather than using an in-house UKRI classification.  **Background to the Requirement**  UKRI has experienced significant challenges in managing and classifying its portfolio to structured classification systems This is required to understand where we are investing our funding and to use the analysis to inform strategic decision making. Manual processes are time consuming, error prone (individual bias) and inefficient.  UKRI purchased the current system in 2018 to provide auto-classification to structured classification systems to enable reporting to external agencies and our boards. The current system provides additional functionality that has been used: as a linked research information system; bringing together publications, awarded grant applications from ~400 global funders, policy documents, patents, clinical trials, publication citations, enabling users to explore the connections between them and delivering an autoclassification and search function.  Our current contract ends 31st March 2025.  There is a need to provide a solution from 1st April 2025 so that UKRI statutory reporting can continue. There is a reputational risk should this not be in place.  **Scope**  The acquisition and implementation of an AI/ML based auto-classification for structuring unstructured data into predefined categories. The scope includes the provision of software, implementation, integration and training services to ensure the solution meet requirements.  In Scope  AI/ML solution: the development of an AI/ML model to classify unstructured data into a structured system.  Data integration: Integration of the classification system with existing data sources  Model training and customisation: the vendor is responsible to train the AI/ML model using existing or new datasets to ensure classification accuracy.  System deployment and testing: deployment of the solution in a live environment, including user acceptance testing to ensure it meets requirements.  Post deployment support: ongoing support, maintenance, periodic retraining of the AI/ML model to maintain performance over time.  Out of Scope  Provision of hardware or IT infrastructure  Custom development beyond core classification functionality  Potential demand: demand may fluctuate, based on the volume of new applications being processed. The vendor should ensure the solution can scale to accommodate this.  Known constraints: compliance with data protection and our Data Protection Privacy Impact Assessment and meeting GDPR requirements.  Period of agreement: the agreement will be in place from 1st April 2025 for a year.  **Requirement**  The vendor must provide an AI/ML based system capable of classifying unstructured data (text) into a predefined structured classification system.  The system should support real-time or batch classification, depending on the use case.  The solution should use relevant machine learning algorithms (e.g., supervised or unsupervised learning) and should provide high accuracy and efficiency.  Output structure: The structured classification systems that need to be created are: Sustainable Development Goals (SDG); OECD sector codes; Health Research Classification System (HRCS); Research, Condition and Disease Categorization (RCDC); Fields of Research (FoR); Common Scientific Outline and Cancer Type Codes; Units of Assessment (UoA).  The solution must integrate with our existing IT infrastructure. It must be able to securely move data from our systems, and provision the output so it can be made available in UKRI’s databank, enabling users to view and query.  The vendor should have proven experience in developing AI/ML solutions for classification particularly in the output structure described above. The provided solution must have the functionality to deliver these structured classifications from the start of the contract.  The vendor must ensure that the solution complies with data protection regulation (e.g., GDPR) and implement necessary security measures to protect our sensitive data.  The vendor should be able to demonstrate how well their system works with our data and its ability to meet our classification needs. |

*This form is an expression of interest only to understand if and what solutions are available that will fully meet the contracting authority’s needs; the full details of the project and the desired outcomes will be provided should a future tender be advertised. At this stage there is no commitment that a tender exercise will follow.*

# **Procurement Plan / Dates and Budget**

|  |
| --- |
| **Project budget** – £690,273.00 excluding VAT over the full life of the contract.    **Indicative advertising timescales for a further competition** – February 2025    **Indicative tender live period** – 5 weeks    **Indicative Contract start date** – 1st April 2025    **Contract duration** – 1 Year + 1 Year (optional extension) |

# **Validation Questions**

|  |
| --- |
| To ensure that this potential procurement maximises appropriate bidder responses we have the following questions that we would like to pose to interested suppliers:   1. **Supplier Capability to meet the specification**– Please can you confirm if this project is something that your organisation would have the capability to undertake? Has your organisation undertaken similar projects in the past? |

# **Response from Supplier:**

|  |
| --- |
| *Based on the project detail provided, would your organisation be able to bid based on the requirements outlined?*  Yes / No  *Any further comments* |

Please provide a response to this question by 16:00 on Friday 15th February 2025 by emailing this to [ddatprocurement@uksbs.co.uk](mailto:ddatprocurement@uksbs.co.uk)

|  |  |
| --- | --- |
| Name: |  |
| Email: |  |
| Organisation: |  |