THE NATIONAL ARCHIVES

REPLACEMENT HISTORICAL MANUSCRIPTS COMMISSION DATABASE

DEADLINE FOR TENDER SUBMISSIONS – 12 noon (UK TIME), 25 November 2019

CLARIFICATION QUESTIONS AND RESPONSES

The National Archives has received a number of clarification questions relating to this opportunity. Those questions, and their associated responses, are detailed below.

Q1: *Can you provide a link to each service separately? If they are integrated as a module can you please point us to the module at the TNA browser?*
A1: This is not possible as the service is internal to our organisation

Q2: *What are the indexes exactly?*
A2: The indexes are generally high level descriptions of collections. So an example index entry might be "Agricultural Committee: Minute Books" or it might be "Jessie Boot (1850-1931): collected papers". But generically by 'index' we mean 'entity'. E.g. The Manorial Documents Register tables constitute an index of manorial documents and manors.

Q3: *Do we need to support upload of data (files)? If yes, what kind of files?*
A3: Yes – these are essentially .csv flat files of similar. Currently we use Microsoft SSIS to validate and load the files into SQL Server, similar loading processes will be required, use of other similar technology is acceptable.

Q4: *Do we need to be able to visualize the data?*
A4: Visualisation is a nice-to-have but is not strictly necessary as part of this tender. We currently use MS SQL Server Reporting Services, and plan to use MS Power BI. If your proposal includes visualisation, please detail this in your submission.

Q5: *What kind of reports do you plan to run?*
A5: We currently use MS SQL Server Reporting Services and plan to use MS Power BI. Also data is extracted for use in Discovery by SQL statements which load data to the Mongo database. We currently obtain basic reports such as 'how many entities in the system' (how many manors? How many archives? How many entries edited last month?). We would like to be able to run more sophisticated queries in the future ('where in the UK are collections about scientists concentrated?').

Q6: *Do we need to provide payment service or only to calculate the prices?*
A6: No, there is no need to provide any payment service. The system should calculate totals which we will pass internally to Finance.

Q7: *From which sources you want to ingest data?*
A7: From Excel spreadsheets and MS Access databases.

Q8: *Bulk editing is editing of the files (Excel spreadsheet) uploaded to the browser?*
A8: This would be ideal, however a client/server approach with software installed on a Windows 10 device is also acceptable.

Q9: *From which databases the browser should be able to upload data?*
A9: From Excel spreadsheets and MS Access databases.

Q10: *Can you tell us more about "Provision of CRUD (Create, Read, Update, Delete) operations in the new interface, using a set of screens based on current views of the data"?*
A10: This refers to normal data maintenance tasks creating and editing existing data in the database.

Q11: *What do you mean by "production of bulk emails"?*
A11: Essentially production of standard awards or rejection messages by mail merge to the responding archives. The list is 250-300 repositories.

Q12: *In point 3.1 [of the Invitation to Tender] – could you clarify what is the discovery infrastructure, what is the current data structure and how does it interact with the discovery infrastructure currently.*
A12: Discovery is an application based on the mongo database. TNA populate the Mongo structure from an intermediate SQL Server database which pulls together data feeds from several sources inc HMC Admin. Assuming the proposed solution will provide the HMC feeds to that SQL database, TNA will operate that interface.

Q13: *What is the current volume of data and what is the expected incremental volume expected totally for your storage volume.*
A13: Current volume is approx. 7GB for the HMC database, for your proposal please assume growth of 10% p.a.

Q14: *The tender mentions a project completion date of 31/3/2020. Is this a firm date?*
A14: Strictly speaking, this is a firm date apart from minor issue resolution. However in Section 3.1(a) of the ITT where we have talked about an agile approach. Potential suppliers should, as far as possible, propose what is reasonably deliverable within the timescale and budget.

Q15: *Please note that [Section] 1.3 [of the Invitation to Tender] mentions a grant management system. Could you confirm that this is a mistake?*
A15: Apologies, it is not directly relevant – the final line should have been omitted. We manage grants and we have a grants management system. That is nothing directly to do with this tender.

Q16: *How many staff will access the system?*
A16: Potentially up to 65 users. In practice, the system has about 10 intensive users and the number of simultaneous users scarcely ever rises above 5.

Q17: *How many third parties will access the system?*
A17: None.

Q18: *Could you give examples of bulk ingest capabilities you need, outwith the ascensions/new burdens process?*
A18: The main other example from our processes as they are today would be the ingestion of Access databases of manor surveys carried out by project officers in the counties. These

describe the records they have identified and are added to the Manorial Documents Register tables in the database at the rate of about 2-3 a year.

Q19: *Could you provide examples of bulk edit capabilities required?*
A19: A simple query based example would be “return me all the records which contain the word “Surrrey” [sic, note typographical error]. Replace all instances with “Surrey”. Or return all URLs beginning ‘http://www.durhamrecordoffice.org.uk/…’ and replace them with the single URL ‘https://www.durhamarchives.org.uk’ (alternatively swap out the prefixes).

Q20: *Is documentation available on the structure and content of the database?*
A20: This will be shared with the appointed supplier.

Q21: *Could you provide a list of interaction points between this system and other systems, both internal (eg Discovery) and external?*
A21: Data load from Excel and Access via MS SSIS, extract through SQL statement in customer written programs to populate the Mongo database underlying the Discovery service.

Q22: *How large is the database? Ideally number of tables and number of records and size in GB per table, or alternatively total size in GB.*
A22: Approx. 7GB.

Q23: *At what rate will the data grow on an annual basis?*
A23: For the purposes of your submission, please work on an assumption of 10% annual growth as a maximum.

Q24: *Will a proposal be considered if we are in the process of obtaining Security certification?*
A24: The contract will require security certification of an agreed standard to be in place before signature. A bidder could be selected as preferred supplier, but no contract will be signed without certification. Preferred status will be revoked if the supplier does not achieve certification within an agreed, reasonable timescale

Q25: *Will the new HMC Admin system be used by TNA staff internally only? How many concurrent Users are expected to be using the system during a peak period please?*
A25: Yes. We envisage a maximum of 15 concurrent users.

Q26: *Please share the API Interface specification referred in para 3.2e [of the Invitation to Tender].*
A26: The data to be loaded to the TNA Discovery (Mongo based) service is extracted into a SQL Server database called EAV, from several sources, one of which is the HMC service.
The relevant sql queries are attached separately. The process runs as follows, firstly there is a daily population of the Audit tables in EAV, these are AuditUMArchon, AuditUMFA and AuditUMNRA which just capture the relevant Audits from the HMC database Audit table in relation to the 'SourceID' for the type of audit, ie. SubjectID for FA, RepositoryID for Archon and ArchiveRecordID for NRA with the original table and AuditType. This process is run by the job UpdateMongoHMC Populate AuditUM tables which runs the procedure UpdateMongoPopulateAuditUMTables. This is the control procedure for the process which defines which tables and the type of updates to capture. It deletes existing records for the input date, captures the audits and deletes any duplicate updates for and 'SourceID'. For example Repository updates:
Job UpdateMongoHMC Populate AuditUM tables Runs
The job runs SP UpdateMongoPopulateAuditUMTables @AuditDate
The SP runs SP UMArchonGetRepositoryUpdatesForTable
The SP runs SP UpdateMongoArchonGetAudit which selects from a view FROM [dbo].[v_Audit] and the view selects from the HMC linked Server

Q27: *Please share the Interface specification with Discovery platform.*
A27: Assuming the proposed solution will provide the HMC feeds to that SQL database, TNA will operate that interface.

Q28: *Has TNA considered the different licensing models for SQL Server and the costs for SQL Server Reporting Services, SQL Server Integration Services?*
A28: Yes.

Q29: *Is TNA eligible for reduced licensing costs, like the NHS?*
A29: Yes

Q30: *What current licenses do you have SQL Server now, and what version(s), how many cores?*
A30: Current versions for SQL Server 2012 and 2016 standard edition on various servers from 4 to 8 cores

Q31: *Seems to be listed as ISO200071 in your document (Extra zero)? ISO20071 is more about the ecosystem of an enterprise in terms of maintain Information Security standards. For example, an organisation with one secure application would not pass ISO20071 if other systems weren't meeting the standard? Web based systems can use OWASP to try and improve security.*
A31: TNA currently holds ISO27001 certification (apologies for the typographical error). If your proposed solution is a cloud service, please specify in your submission what level of security assurance is in place. Our reference point in this area is Cyber Essentials Plus

Q32: *Post go-live, the project would enter a support/maintenance phase. Is TNA committing to some spend for this support?*
A32: Please detail your proposed support process, with price(s) as appropriate.

Q33: Do you have ideas on numbers of users who are data controllers, viewers, and reporting analysts?
A33: Relatively small numbers use the service, as it is used as a feeder system. Estimated is approx 5 people use the system at any one time and that there would be 65 separate IDs.

Q34: *How many data sources are there?*
A34: 6

Q35: *Is it the case that there are only a few data sources but multiple instances of them? For example, there may be twenty instances of one spreadsheet?*
A35: Yes

Q36: *How stable are these data sources? For example, spreadsheets and access databases are notorious for record locking issues.*
A36: These issues occur occasionally and are resolved.

Q37: *How well are these sources validated? For example, spreadsheets are very easy to "break" by adding different data types.*
A37: Current process include data validation - occasionally we have to unload and reload data.

Q38: *What support issues do you encounter now, when uploading these data sources?*
A38: Occasional problems with duplicate records or non-loadable characters – these issues are normally identified as part of the pre-load data assessment.