

Does this style guide exist? It says this will be created, as part of the project or is it out of scope of the project? Annex A - The website must conform to an approved style guide.

No, it does not currently exist and will need to be created separately as part of the overall project. We envisage the creation of the style guide to be a component of the optional requirement described in Section 8 of the ITT

Does this mean we need to provide hosting? Annex A - The website must be available and measured against specific SLA targets.

The Supplier is only expected to provide a build and test environment for the website. The live environment will be externally hosted by a company (Etisalat) in the UAE.

Arabic search can be different from English search. For example if I search in Arabic for a word in singular, the word in plural is not as it is in English with the same character starting, it is a total different word. Does phonetic search need to be provided? Annex A - Search in Arabic.

We are keen for potential suppliers to propose solutions in their tender response to any challenges of this type that they have identified.

Advanced search can mean a lot, like entering date range, select out of list of authors or origin. Can we specify the values of advanced search? Annex A - There must be an advanced search option for historic document searches.

The advanced search options have not been defined at this stage. We expect these to be fixed during the more detailed scoping phase of the project with the chosen Supplier. The final decisions will be based on a mix of user needs, the nature of the content to be searched across and the technical complexity required to deliver these.

Will the documents be OCR'd before they arrive on the website, or is OCR part of the solution? Annex A - OCR search.

Yes, they will. OCR is out of scope for this tender.

Does the search result display both languages, or only one? And should the user have the ability to choose in which language to search. e.g. using an English word to search for a document in Arabic. Annex A - OCR search.

It is currently anticipated that search will query two sets of data (for searches within the document content). The first set of data is the catalogue description that will exist for each catalogue entry and this will exist in both English and Arabic. The second is the OCR data which will provide a transcription of the actual text from within documents. This will only be in English as the source documents are primarily English (even where an Arabic document exists, there is usually an accompanying English translation within the file). It is currently anticipated that searches done in

English will provide results in English and searches done in Arabic will provide results in Arabic. There is no fixed requirement to translate search terms between languages although this might be considered following an evaluation of the technical complexity, cost and if it could be delivered within the project timeline.

Where are these notes added? Will there be a user page? Annex A - Category 3 – User Interaction.

We are keen for potential suppliers to propose solutions in their tender response to any challenges of this type that they have identified.

**User preferences for search, what exactly are user preferences, can you give us an example?
Annex A - The search results page must be user configurable.**

To be defined, but could include things like number of searches to show per page, filtering by pre-defined categories, whether or not to show results from OCR searches. We are keen for potential suppliers to propose solutions in their tender response to any challenges of this type that they have identified.

We assume notes are personal and are not public; is this correct? Annex A - User has the ability to add notes.

The current thinking is that they could be either and the final solution may depend on the design/technical considerations. Having notes that could be seen publically might be a good way to generate debate/dialogue between users. Private ones would be good for individual referencing.

We assume that comments are public, correct? Annex A - User has the ability to add comments.

Yes, the idea is that users could share views on articles published and reference other content that is relevant and might be interesting for others.

**Where is the e-newsletter generated, where are the email subscribers managed, do we need to forward them to an email marketing tool, if yes which one is it? Or is this part of the software?
Annex A - User has the ability to request e-newsletter updates.**

The creation and management of e-newsletters is out of scope for the initial phase of the project. It is assumed that any email addresses that are captured by the website can be exported into an independent system that will administer these and generate/send emails.

Do they need to be watermarked before downloading is allowed? Do users need to be known, logged in before they can download? Annex A - Category 4 - Historic document content must be downloadable from the website.

No, there is no requirement for images to be watermarked. It has not been finalised if a log-in will be required to enable a download to commence. The final decision may also be influenced by the proposed technical solution and if/what login looks like (and how it works).

Depending on the format this can be achieved, for example if we have a document (PDF) zooming and other document related functions are there. However brightness and contrast are image format features, can we assume that functions relate to their nature of format? Annex A -

Category 4 - Any historic document content "viewer" must have a zoom function and other standard viewer functions (e.g. rotate, contrast, brightness, invert etc.).

All images will be provided as 300ppi resolution colour images. The imaging process tries to capture the documents "as they are" and there is not a processing step involved in their creation to enhance the image to make it easier to read. It is expected that whatever online image viewer is created has some limited functionality to enable users to alter individual images on the fly based on their requirements.

How does the user submit this content, over a web based form or via email ? Who are these users, registered users or public users? Annex A - Category 5 - Ability to review any user generated content and approve it for publication to the website.

This is currently not defined and we have no specific solution in mind. It will be up to the Supplier to propose an approach (ideally based on experience and/or standard market approaches that have been proven to work).

Does this mean to support the automatic ingestion of this formatted catalogue? Annex A - Category 6 - The website must have catalogue administrative system that ingests, manages digital images, catalogue descriptions and technical metadata (assets).

The site will launch with approximately 500,000 images. These will be arranged in a pre-determined hierarchy that mirrors TNA's own catalogue structure. The admin system needs to have a mechanism to be able to ingest the images, their structure and relationship to each other and all of the associated metadata (catalogue descriptions, translations etc.) as part of an automated process. We cannot load this information manually record by record, image by image. It may be that the initial load is completed by the Supplier as a manual bulk load but there needs to be some form of automated system in place eventually to enable new content to be loaded easily post launch.

Now we know where the hosting will be. Can we ensure to have full access to the system from the hosting provider? And who will it be? Annex A - Category 6 - The system will require interoperability that will enable the underlying digital storage of assets to be hosted by a cloud services supplier based in the United Arab Emirates.

The hosting supplier for the live website will be Etisalat, based within the UAE. It is expected that full access to the service will be allowed in order for the installation of the website and for final testing etc.

Does this mean the XLS needs to be produced from the system or manually generated and uploaded to the DAM? Catalogue Descriptions (XLS).

The catalogue descriptions will be generated as a separate work stream within the project. The output format of these descriptions is .xls but potentially could be converted to allow ingest to match system requirements. There is no requirement for the system to generate this content itself.

Does this mean the CSV needs to be produced from the system or manually generated and uploaded to the DAM? Cataloguing Metadata (CSV).

The CSV will be generated externally to be loaded into the DAM. There is no requirement for the system to produce it (although there is a requirement that there could be a bulk export of data from the system, but this has not been scoped out in terms of output format).

Will it be produced or generated out of the system? Bi-Lingual Catalogue Descriptions

All catalogue descriptions and their translations will be generated outside the system and then imported.

Is this a list of platforms or which one is preferred? WEB, APPLICATION AND DATA CENTRE HOSTING.

The list of platforms was provided by Etisalat, the hosting company for the live website as ones they will host and support. There is no preference which is used by the Supplier as long as it is one from this list.

Are there any site visitor numbers, expected so we can size the platform? Architecture.

This is a new service so we have no historic data to base against. We have not yet scoped the likely levels of traffic so cannot make any predictions. Suppliers may want to suggest a scalable solution that can be ramped up or down depending on demand (for example, there may well be a high level of traffic at launch, but this might then subside for a period).

It seems like Design and Layout needs to be done during the project? Design – CI.

Yes, this is a new platform and a new brand. Design and layout are expected to be produced as part of the brief of the Supplier. We envisage the creation of a style guide to be a component of the optional requirement described in Section 8 of the ITT.

Is there an existing Collection/DAM tool / system in place within the National Archives that has been used for any steps in the existing content workflows described in Annex B?

No, each process is operating independently at the moment although we are maintaining project oversight and tracking via our Project Manager. TNA does not have general content workflow tools to automate this.

Are high-resolution originals expected to be made available via the public-facing website?

This has yet to be finally defined. TNA's approach on our own website is to provide a lower resolution "display" version of the image for download, however there has been discussion about allowing certain users (e.g. identified and flagged academic users) to be able to download higher resolution files. The final decision will be based on an evaluation of user needs, technical options and complexity and business risk appetite.

Is the Gulf Archive expected to be expanded on over time? If so, what would be the estimated rate of change?

Yes, there are expected to be future phases that add new content over time. This could be of the scale of 500,000 additional images (and associated metadata, OCR output and catalogue descriptions) per year. TNA likely has enough source content to maintain this level of growth for at

least 5 years post launch, however anything beyond December 2018 would fall under separate contractual agreements. It is also anticipated that future phases will include elements of web development to add new functionality and/or improve the existing experience.

Does the hosting infrastructure have to reside in the UK or the UAE – or anywhere else?

The live site will be hosted in the UAE by Etisalat. The development and test environments must be provided by the Supplier as part of the service to create the website. There are no restrictions for where the infrastructure needs to reside (all the records that will be used to populate the website are open public records and freely accessible at TNA).

Are there any restrictions on document access? (i.e. sensitive data, only specific user groups, countries etc.)

No, all the records that will be used to populate the website are open public records that are currently available at TNA. It is possible that the NA, UAE may decide to not publish specific documents on the live site (reflecting local political sensitivities) and the system will have to accommodate this functionality, however the records will continue to be available via TNA at all times.

Is the supplier-customer relationship with the UK National Archives (UK NA) or The National Archives of the United Arab Emirates(UAE NA)?

The contract to build the website will be between the UK National Archives and the Supplier. Following launch, ownership of the finished site will pass to NA, UAE for ongoing business-as-usual management.

Is the procurement and on-going leasing of hosting infrastructure the responsibility of the supplier or UK NA?

The hosting of the development and test environments will be the responsibility of the Supplier, operating under contract to TNA (UK). The live site will be hosted by Etisalat in the UAE under contract with the NA, UAE. It is expected that any ongoing support and maintenance, post launch will be agreed and contracted directly with NA, UAE.

If the data/records for the platform are supplied by the UK NA (ITT, Sec 1.4- primary source material from within TNA), what would any on-going support & maintenance & content requirements be post-handover to the UAE NA?

It is expected that any ongoing support and maintenance, post launch will be agreed and contracted directly with NA, UAE.

Will any ongoing services or agreements post-handover be with the UK NA or the UAE NA?

It is expected that any ongoing support and maintenance, post launch will be agreed and contracted directly with NA, UAE.

What are the UK NA defined standards & project constraints mentioned in the ITT, sec 2.2?(is it the content workstream in Annex B?)

The standards are defined in Annex B (mainly around the technical infrastructure) and the business constraints are in section 3 of the ITT document.

Are there any traffic estimates, and any defined SLA's referring to Annex A, Cat 1 (SLA Targets & substantial traffic)?

This is a new service so we have no historic data to base against. We have not yet scoped the likely levels of traffic so cannot make any predictions. Suppliers should suggest SLA targets based on their ability to meet these targets and on demonstrated industry best practice.

There is no confidentiality statement associated with the tender, is this correct?

Yes.

For the avoidance of doubt, is TNA adamant that the hosting provision has to be cloud-based with datacentres based in the UAE? Is there room to work with you in identifying the best performing and most cost effective solution?

Yes, this is a base requirement of the NA, UAE who will be the eventual operators of the live service.

Regarding the 'Web, Application and Datacentre hosting', does the general overview list confirmed requirements or representative examples? Is there some flexibility here?

This list was provided by Etisalat (the UAE hosting company) when asked what they would support so submissions should be based on this list. There may be some flexibility during the development phase based on a dialogue with them if an alternative has demonstrable benefits to the project.

In what format will metadata be supplied? Will the organisation of metadata be consistent across all assets? Would you be able to supply us with sample metadata?

The details of the expected outputs from the various work streams within the project are described in Appendix B of the ITT as we currently expect them. The only one not yet specified is the output of the translated descriptions as we have not yet engaged the supplier to complete this work (and so have not finalised the requirement). We expect the application to be consistent. We do not have sample metadata at this stage as it is still in creation or has not yet begun. Test materials will be supplied as required once the project has been initiated and a timetable agreed.

Do you have specific SLA targets and metrics in mind?

Suppliers should propose SLA targets based on their ability to meet these targets and on demonstrated industry best practice.

Can you confirm 'catalogue descriptions' refers to finding aids and not to asset metadata?

Catalogue descriptions refers to descriptive metadata that will be associated to images within the collections at the appropriate hierarchy level (i.e. a single catalogue description will cover all images that make up a specific catalogue entry e.g. FO 371/1 where FO is the Department, 371 the series and 1 is the piece). We define finding aids as specific pieces of static content that might be published on the website as a separate document that users can read to gain insight into how to research the

collections (e.g. a guide to understanding how the Foreign Office catalogued records in the early 20th Century).

Would you need the platform to support view of 300ppi tiff master files or is it acceptable to work with jpeg/jpeg2 digital surrogates?

This has yet to be finally defined. TNA's approach on our own website is to provide a lower resolution "display" version of the image for download, however there has been discussion about allowing certain users (e.g. identified and flagged academic users) to be able to download higher resolution files. The final decision will be based on an evaluation of user needs, technical options and complexity and business risk appetite.

In what format would OCR output be supplied? Our preference would be to support HOCR or other similar outputs which would allow use of co-ordinates to display search results and result in better user experience than through use of searchable PDF or plain text.

The current specified output for OCR as detailed in Annex 2 is searchable PDF v.2.0 and Plain Text Output. If another format is identified and agreed more suitable (and affordable within the project budget), that could be adopted. Suppliers should make recommendations as part of their submission backed up by evidence and based on their relevant experience if these recommendations differ from what is asked for.

Is there a preference between open-source or proprietary CMS?

No, although Suppliers should specify as part of their submission what they intend to use. There are implications on ongoing IP ownership following completion of the project that need to be fully understood, however our evaluation does not bias in favour of a particular approach in this area.

What does success in regard to this project look like for TNA and NA?

There has to be a functioning website live and accessible by the general public by the prescribed date. This website must provide access to the agreed amount of digitised content and all associated metadata. The site must exist in both English and Arabic versions. The website must contain the agreed functionality.

How will that success be measured/assessed?

The success criteria will be the deliverable components meeting the specified requirements.

What is the vision for the site beyond its launch in 2018?

There are expected to be future phases, subsequent to launch, that add new content over time. This could be of the scale of 500,000 additional images (and associated metadata, OCR output and catalogue descriptions) per year. TNA likely has enough source content to maintain this level of growth for at least 5 years post launch, however anything beyond December 2018 would fall under separate contractual agreements. It is also anticipated that future phases will include elements of web development to add new functionality and/or improve the existing experience.

Could we see some sample content?

At this stage there is no complete sample content to share. Digitisation and cataloguing is ongoing, however the OCR and translation work has not yet begun. Therefore a sample of all the different components that need to be integrated into the website cannot be provided at this time. There are many examples of digitised content available on The National Archives' website that can be downloaded, but these should not be taken as an exact match for what will be produced within this project (the display images on our Discovery system are not the same file format or standard as the master images created during the digitisation process).

Apart from Documents, Maps and Photographs are there plans to include any (all?) of the following: video, audio, 3D objects, any other item types?

There are no plans to include any other types of content for the initial launch phase of the project than those described within Annex B of the ITT. It is likely however, that future phases of work post launch will include other media types (e.g. video, audio). Suppliers may want to include information regarding the support of other format types as part of their response to provide evidence of scalability and future proofing.

Will you provide OCR data in a structured format that includes positional information (e.g. ALTO), or just as plain text and searchable PDFs as you mention.

The contract for completing the OCR work has not yet been awarded so output specification has not been finalised. Our base requirement is laid out in Annex B, however other output types could be considered following a discussion with the appointed Website ITT Supplier. Cost, complexity and impact on the delivery timeline will likely influence the final decision.

Will the data filters be simple descriptive names (e.g. "Abu Dhabi") or contain extra information (e.g. parent country: United Arab Emirates; geocoordinates: N 24°28'00" E 54°22'00"; Geonames ID 292968; etc.)

The data filters for launch will all be simple descriptive names and or topics e.g. Internal Affairs, Finance, Visits.

Do TNA and NA have any preferences or requirements in regard to thesaurii and authorities for names and index terms?

We recognise that there is inconsistency within the archival record regarding how things are described (changing place names over time for example) and the search functionality of the site will need to anticipate this. It is expected that there will be an authority list for identified terms which will be added to over time. The ITT response should demonstrate how the Supplier has dealt with similar challenges in the past.

Will the searchable PDFs that you provide be ready for directly publishing on the website, or will they require further processing before publication?

This will ultimately depend on the final technical solution and also what is agreed will be the "deliverable" for the customer when downloading. Integration of OCR data will also be a factor that needs to be considered. We are looking for suppliers to propose solutions based on their experience.

How is the existing archive data (for the initial set of 500,000 items) managed? Is the cataloguing metadata in an existing TNA system?

All of the images that will be created are from TNA records that have already been catalogued and organised. There is a fixed hierarchy in place and image naming convention will reflect this. The cataloguing data will match within this hierarchy.

How do you anticipate the initial archive data will be submitted to AGDA?

There are a number of possible ways the archive data will be initially loaded into the system and Suppliers should propose their preferred method (based on evidenced experience). For example, all the complete content could be bulk loaded in a single exercise, content could be loaded in batches, batches could be organised by type (e.g. all images, then all OCR data) or combinations. We anticipate the initial loading of content will be completed by the Supplier.

How will content from sources outside TNA be submitted to AGDA?

This is out of scope for the launch phase of the project.

In the catalogue administration system, when you refer to editing content, do you mean replacing an existing record with a new version uploaded as a whole in the formats you mention (CSV, XLS), or actually altering individual fields of the archival metadata via an online editing interface?

Suppliers should articulate how they plan to meet this business requirement based on their experience and expertise. Bids will be evaluated on the quality of the responses.

Can you give more details of the anticipated catalogue hierarchy?

The catalogue hierarchy can be viewed on The National Archives website. One of the specific series that has been identified for inclusion in this initial phase is FO 1016.

<http://discovery.nationalarchives.gov.uk/details/r/C8315>

(For the avoidance of doubt) we assume that all digitised content submitted will be for presentation purposes only, not preservation - i.e. our solution could include a pipeline to transcode source formats and discard input copies. Is that correct, or do we need to present a preservation solution also?

There is no requirement within this ITT to provide a digital preservation solution for the preservation copies of images and metadata.

We have been unable to open the embedded attachments for the Etisalat OneCloud platform. Please could you supply these as separate documents?

Yes, these can be supplied upon request to procurement@nationalarchives.gsi.gov.uk

Are TNA and NA happy to use Google Analytics?

Suppliers should include their proposal for website analytics as part of their submission. Bids will be evaluated on the quality of their response.

How many users are anticipated in the first year and beyond?

This is a new service so we have no historic data to base against. We have not yet scoped the likely levels of traffic so cannot make any predictions. One of the key pieces of work early on in the project will be to do some work on audience groups and this will likely lead to predictions of use based on the estimated sizes of those.

Do TNA or NA have a user testing panel that we can use for testing prototypes and work-in-progress?

Not yet, but work is beginning to establish one. It is likely this will be in place within the first 3 months of 2018.

Comment facility (Annex A, Category 3) - is this only for the 'static' content, i.e. blogs and articles? Or is it envisaged to be also available on archive documents?

Ideally yes. We think it would be a good feature for users to be able to make notes on specific documents as they research them (these notes could be public or private). As with all the requirements, the final outcome and solution will be dependent upon the technical complexity, cost and time required to complete.

It would be helpful to know how the teams at TNA and NA are structured including relevant roles and responsibilities.

The project is being run and managed by a team at TNA on behalf of the NA. Although there will be required dialogue with NA especially in the finalisation of the scope and testing phases (as well as the go live), TNA will take the lead with the Supplier and primarily be responsible for liaison with NA. A TNA project team has been established (including project manager resource, business owner resource plus delivery teams for each of the work streams). Quarterly project steering group meetings are in place (joint with TNA and NA).

It would be helpful to know what the arrangements are for communication between TNA and NA and how the developers should participate.

TNA will take the lead and facilitate all communication with NA. The Supplier will primarily interface with TNA during the project. There are already strong communication processes established with NA by TNA and these will continue (and adapt as required within the project).

How will the sign-off of deliverables work as between TNA and NA?

TNA have established a high level project plan with NA which is subject to formal monthly review. As details of the website build are finalised, milestones will be agreed and added to the project plan. TNA will work with the Supplier to meet these milestones and manage the expectations of NA.

Apart from launch on 2nd December 2018, are there any significant milestones in the TNA's project plan that we should factor into ours?

There are delivery milestones for each of the project work streams, all of which have interrelationships. These milestones are still being established as some of them will be delivered by

third parties on behalf of the project (e.g. translation). For the purposes of the submission, Suppliers should articulate their own milestones for delivery and highlight any deadlines required by others in order to complete (e.g. all translation must be completed and loaded by XXX in order to hit the launch date). The final agreement of all project milestones will be completed during early stages of the detailed project planning.

Are there any planned visits to the UAE that we should consider attending?

There are no dates currently planned, however there is an anticipation that there will need to be some visits over the course of the project. The first is likely to be early 2018 (mid to late Jan/early Feb) to workshop the final scope, discuss project planning and audience elements (plus branding if applicable).

Do TNA/NA have any requirements or preferences in regard to development methodology?

No, Suppliers should specify their preferred approach along with supporting evidence of why that approach is appropriate. References should demonstrate competency and experience of execution using the proposed methodologies.

Do the TNA team have any experience with Agile software development?

Yes, TNA actively use Agile for development of our own systems. The allocated Project Manager for this project is Agile trained.

Does TNA have a template for the contract? Are we free to table our own suggestion?

The contract will be awarded under our standard terms and conditions. Please see section 7 of the invitation to tender for more information and a link to these terms.

Does TNA have a preference in regard to milestones and payment schedules?

No, Suppliers should specify their preferred approach along with supporting evidence of why that approach is appropriate.