**ANNEX B**

**USER GROUPS**

The AGDA Project has been created in order to fulfil the need to:

* Enhance the understanding of cultural and historical depth and breadth of the Arabian Gulf region.
* Increase the availability of this information and make it easily accessible to academics and researchers as well as interested non-specialists and the general public.

A majority of the historical records that relate to the Gulf Region exist in archives around the world. For historians and academic researchers it is therefore a significant challenge to access primary source material, especially if those records have not been digitised. For a non-specialist or a citizen of the Gulf Region who has little or no understanding on how to conduct research, or where to even begin, this becomes even harder.

With the creation of the online digital archive, the NA has the goal to provide this much sought after information to this wider audience. The main end-user audience groups will be researchers and academics but there is also an appetite to make the content available to and understandable by the general public and schools. This creates an interesting challenge as the online digital archive needs to appeal to different groups who have different expectations and skills.

The identified end-user groups are:

**Researchers and Academics** – This end-user group are generally used to doing research. They understand catalogue structures, are familiar with searching through large volumes of information and have (usually) enough background knowledge to be able to identify relevant information quickly. They are used to interpreting primary source material and have an appetite to see the original documents (or digitised versions of).

**General Public** – This end-user group has a passing interest in history and would like to learn more about the Gulf region. However they have little or no understanding of archival practices or research methodologies. They respond well to images (photographs and maps) and would generally get more value from reading an article about a topic than from reading the original source documents themselves. Their expectation of research is the ubiquitous Google search.

**School Children and Education** – One of the long term aims of the project is to create educational resources based around the content that can be used for educational purposes. A clear way to ensure future generations are passionate about their history is to engage with them at an early age and demonstrate that history is accessible, fun and interesting. By providing learning resources that can be used to base school lessons on and getting children to interact with, understand and discover historic records the hope this that this will be realised. School children generally have a high level of computer literacy and are not afraid of technology although they need to be guided in their work and have very little background knowledge with which to apply context.

**CONTENT WORKSTREAM**

The content work-stream of the overall project will be responsible for the creation of all archival and historic material that is presented by the AGDA. The outputs of this work-stream are detailed below and TNA, as lead coordinators of this work-stream will ensure these outputs are fit for purpose and meet the required standards. It is expected that they will be loaded onto the website and thoroughly tested as part of the process to create the AGDA. The activities are:

* In-depth research into relevant and note-worthy historical primary source material for the end-user audience
* The preparation of all primary source material for digital imaging
* The imaging of all identified records
* Quality assessment of digitised images
* The creation of technical metadata to international standards that will provide linkages between images, the catalogue structure and any OCR outputs
* The creation of searchable metadata using Optical Character Recognition (OCR) technology
* The creation of catalogue information that will describe the digitised record and provide context and background
* Identification of suitable data filters and association of these filters to the collections to support comprehensive data searches
* Creation of technical metadata to manage and audit catalogue descriptions over time
* Translation of all descriptive catalogue information into the Arabic language

Below outlines the main activities within the content work-stream. It shows the work-stream’s processes and outputs.

**Digitised Content**

At project completion, the digital archive will contain c.500,000 digitised images. These images will be a combination of:

|  |  |  |
| --- | --- | --- |
| **Record Types** | **Resolution** | **Outputs** |
| Text based documents | 300ppi | TIFFs |
| Large maps | 300ppi | JPEG2000 |
| Photographic prints | 600ppi | Multifile PDFs |
| Large Photographic negatives (e.g. 5x8 inch) | 600ppi |  |
| Small Photographic negatives and slides | 1,200ppi |  |

**Technical Metadata**

A variety of technical metadata specific to the creation of digitised records will be captured for each record. This will assist in process management such as image description and quality assurance, together with defining the provenance of the digital image allowing the project to demonstrate a record’s authenticity and integrity. This will be an input into the technical work-stream.

**Optical Character Recognition (OCR) Content**

All text and some handwritten digitised records will be enhanced using OCR technology. This will constitute a majority of all digitised records. The end-user will be able to access online content through full-text OCR data searches as well as via other means. All OCR-enhanced content will be inputs into the technical work-stream.

|  |  |  |
| --- | --- | --- |
| **OCR Content** | **Resolution** | **Outputs** |
| Text based documents | 300ppi | Searchable PDF v2.0  Plain Text Output |

**Catalogue Descriptions**

Each digitised record will be linked to an enhanced ‘catalogue description’. This is descriptive information (data) which will offer the end-user an explanation of the record as well as help to make the record accessible. Please note that several digitised records can be linked the same enhanced description as many records are sourced from the same volume or series of records.

|  |  |  |
| --- | --- | --- |
| **No of Enhanced Descriptions** | **Size of Description** | **Outputs** |
| c.5000 | c.150 words | .xls |

**Catalogue Description Subject Filters**

Appropriate and note-worthy data subject filters have been identified and will be tagged within all catalogue descriptions. These data filters will also form inputs into the technical work-stream.

**Cataloguing Metadata**

Catalogue metadata will be provided for each catalogue description. This will describe the digitised record and enable support a link between the record and its unique identifier. It will also adhere to international standards for catalogue data modelling.

Each enhanced catalogue description will contain the following fields:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Department** | **Series** | **Piece** | **Images** | **Title** | **Date** | **Description** | **Outputs** |
|  | | | | | | | .csv |

An example of an enhanced catalogue description is shown below:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **CATALOGUING SAMPLE** | | | | | | |
| **Department** | **Series** | **Piece** | **Images** | **Title** | **Date** | **Description** |
| PREM | 19 | 3080 | xxxx-xxxx | MIDDLE EAST. Situation in the Middle East: Iraqi invasion of Kuwait; part 27. | 1990 Sep 01 - 1990 Sep 16 | MIDDLE EAST.  Situation in the Middle East: Iraqi invasion of Kuwait.  This file covers the conflict between Iraq and Kuwait and the position adopted by the British governments and other countries in the region. It contains a summary of British nationals detained in Iraq and Kuwait; correspondence between Margaret Thatcher, the Amir of Bahrein, the President of the UAE, Mikhail Gorbachev, the President of Turkey and the British Ambassador in Kuwait; briefing and notes ahead of meetings relating to the Gulf crisis and the record of a meeting between Margaret Thatcher and the Crown Prince of Kuwait on 3 September 1990.  It also contains correspondence relating to the discussion, enforcement and co-ordination of sanctions, and burden sharing in the Gulf.  It also contains an account of the situation in European embassies in Kuwait. |

**Bi-Lingual Catalogue Descriptions**

The digital archive will enable the end-user with the option to access all catalogue descriptions in both English and Arabic languages. Catalogue descriptions will be translated into Arabic language via an external agency before they are presented to for inclusion into the technical work-stream.

|  |  |  |
| --- | --- | --- |
| **No of Translated Catalogue Descriptions** | **Size of Description** | **Outputs** |
| c.5000 | TBD | TBD |

Please note that OCR-enhanced data will be presented to the end-user in the language of its source document.

**WEB, APPLICATION AND DATACENTRE HOSTING**

The online digital archive will be hosted using cloud services based in the United Arab Emirates. These services will be procured and managed by NA.

The Cloud hosting offering will comprise of **infrastructure and platform services**. These services will be tailored to meet the needs of the online digital archive, however a general overview of each service is outlined below.

|  |  |
| --- | --- |
| **INFRASTRUCTURE** | **PLATFORM** |
| Servers | Windows Server 2012R2 |
| Flexible storage options | Windows Server 2008R2 |
| Capacity | Red Hat Enterprise Linux 6.5+ (32bits) |
| Performance | Red Hat Enterprise Linux 6.5+ (64bits) |
| Network power & cooling | CentOS 6.5+ (32bits) |
| Managed services | CentOS 6.5+ (64bits) |
| 24/7 infrastructure and operating system support |  |
| Physical access control to the hosting environment |  |
| Back-up and disaster recover/continuity services |  |
| Dedicated or shared hosting facilities |  |

An example of sized infrastructure that could be provisioned is shown below.

|  |
| --- |
| 4 VM Servers |
| 6 Cores per Server |
| 96 GB CPU (32-32 GB for front/live servers and 16-16GB for DB and Application/CMS Servers) |
| 300 GB OS Drive each Server |
| 10 TB Storage mapped to any VM |
| Windows 2016 Operating System |
| HTTPS Web service |

There are no technical restrictions for any development tools, operating systems, database management system or any other licensed software that the appointed supplier may wish to use. **However, please note that support at the software level will not be provided. The hosting provider will provide services in the ‘live’ environment only but will support the supplier to migrate the solution from their development / test environment to the live environment.**

A detailed outline of the probable infrastructure and platform services can be found in the two ‘pdf’ attachments below.

 

**TIMELINE**

Below outlines the timeline for the key elements within the AGDA project.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **2017** | | | | | **2018** | | | | | | | | | | | |
| **Key Tasks** | **Aug** | **Sept** | **Oct** | **Nov** | **Dec** | **Jan** | **Feb** | **Mar** | **Apr** | **May** | **Jun** | **Jul** | **Aug** | **Sep** | **Oct** | **Nov** | **Dec** |
| **Record Digitisation Activities** |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| **Optical Character Recognition Activities** |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| **Enhanced Catalogue Description Activities** |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| **Catalogue Translation Services** |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| **Design, Develop Test & Assure Digital Archive** |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| **Go Live** |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |