

RFQQ 18-07
Amendment 1
Complete Set of Questions and Answers

1. Questions from applicant

For this requirement below would you consider our response as acceptable if we are in the process of receiving a UBI number?

The Vendor must be licensed to do business in the state of Washington. The Vendor must meet all state requirements to do business with the state of Washington and have a current Unified Business Identifier (UBI) number.

We understand the current collection is 402,000 pages with a growth rate of 100,000 pages per year. What is the total number of records? Or an average number of pages per record?

Yes, as long as the UBI number has been received before any potential work begins.

The software system we use to generate metadata creates a TIFF, JP2, PDF & XML file for each newspaper page. The web portal should allow the viewer to download a PDF or clip of the page, with zoom in/out capabilities. OSOS will store these files for long term preservation & disaster recovery of the original asset. The vendor will only need to host the files needed for website access. Example: <https://chroniclingamerica.loc.gov/lccn/sn88085189/1898-02-17/ed-1/seq-1/>

Also our growth rate is typically 50,000 pages per year with the potential for 100,000 pages as we continue to build the program.

We understand the current collection is 402,000 pages with a growth rate of 100,000 pages per year. What is the total number of records? Or an average number of pages per record?

Each newspaper issue ranges from 4-16 pages on average, depending on the year. More recent issues tend to have more pages than those from the turn of the century. Our metadata software generates three additional files per raw master file (TIFF 6.0) – a JP2, PDF and an XML file containing the METS/MODS metadata and OCR data. Web users should be able to download a copy of the PDF or they can request a copy of the full TIFF image from us directly. The vendor may wish to use the TIFF or the JP2 or both for the web display.

2. Questions from applicant

#Question 1: In section 1.2, under the "Required deliverables" --- "Software compatibility with Microsoft development platforms"

* Is the term "software" here referring to the web portal software for presenting the Washington Digital Newspaper collection?

* Is this a mandatory requirement for this tender?

* Can you please state the importance/reasons for having the software compatible with Microsoft development platforms?

Clarification on software compatibility: We are currently experiencing a compatibility problem with the limited capacity of our SQL database which holds the OCR metadata and compatibility

of the open source viewing tool provided by the Library of Congress. We wish to achieve the functionality and UX of the open source tool to display a set of pages within an issue. We will consider other options if they are compatible with or successfully independent of this system to display the pages with full-text and advanced search features.

#Questions 2: In section 1.2, under the "Optional deliverables" --- "Provide structural and page/article level OCR metadata in METS/ALTO XML format according to NDNP standards"

* As stated in the introduction of section 1.2, METS/ALO will be provided to us as the source input. Can you please explain how and why METS/ALTO is also expected as deliverables?

Clarification of METS/ALTO standards: During our participation as an awardee the NDNP grant, we are required by Library of Congress to use the METS/ALTO standard for newspaper metadata which allows us to provide full text and advanced search for users at the page level. It also describes the structure of a newspaper issue as a set, ordered number of pages, which are searchable as individual articles or full pages. A user is then able to search a term and find the result highlighted on the page. Our metadata application software automatically provides this so it not required for a vendor to supply this.

We wish to know if there are potential vendors who create or utilize existing METS/ALTO metadata to maintain a newspaper's issue structure and utilize a page or article level OCR process in their end user search tools and display of newspaper page images.

#Question 3: In section 1.2, under the "Optional deliverables" --- "Allow library card access to secure or copyrighted collections"

* How important do you think this requirement is to your collection?

* How do you envision this to work, e.g. can you please provide a use-case scenario for using library cards to access the collection?

This is an optional deliverable to determine the availability of access through a firewall. Many of our newspapers are under copyright protection of the newspaper publisher. We will not be allowed to freely display any pages from the publisher without express consent. As a state library we have many copyrighted newspapers available for research to customers with a Washington State Library card. By signing in through a "paywall" of sorts they would be able to access digital newspapers with a copyright disclosure notice. Newspapers that are copyright free (generally pre-1923) would be available without sign in or library card.

We are reviewing how to make many of our born digital newspaper titles available to the public with respect to copyright law.

3. Questions from applicant

** Under section 3.3.A ("Processing Operation") of the RFQQ there are a number of questions that appear to be related to the process of preparing new newspaper data to be loaded into the web portal. Our understanding though is that the RFQQ is primarily about the process of creating, supporting, and

hosting the web portal, into which the existing 402,000 newspaper pages (i.e. NDNP-compliant files that WSL already has) will be loaded. Then, in addition to those existing newspapers it is likely that 50,000 - 100,000 new newspaper pages will be loaded each year. Those additional pages would presumably be processed to the same NDNP-compliant standard by the WSL and/or a vendor.

We also note that that one of the “optional deliverables” in the RFQQ is the ability to produce NDNP-compliant METS/ALTO data.

The vendor would not be required to produce the METS/ALTO metadata. We would like to know if the vendor would utilize the existing METS/ALTO data generated by our metadata application process.

From that then our assumptions are as follows:

(a) That Section 3.3.A (“Processing Operation”) primarily applies to vendors who can provide the optional deliverable of METS/ALTO data processing, in addition to being able to create the web portal.

(b) That the WSL hopes to be able to work with the successful vendor from this RFQQ process to not only create the web portal, but also to help process some/all of the expected 50,000 - 100,000 new newspaper pages per year in future.

Can you confirm if our understanding as described above is correct? Can you provide any further clarification on this subject?

The vendor would not need to support metadata processing or supply metadata in any way. The vendor would need to create the OSOS branded web portal with advanced full-text search features to access the digital newspaper collection. The successful vendor should have a smooth process for the upload of up to 50,000-100,000 page images per year in small batches prepared at agreed intervals throughout the year.

4. Questions from applicant

1. Aside from the storage of Print Media (Newspaper), does the Office envision itself (i) posting/storing digital content (such as those in the existing archives: <https://www.digitalarchives.wa.gov/>) or press releases (<https://www.sos.wa.gov/legacy/who-are-we/press.aspx>)?

We have a very limited amount of space for print media and continue to focus our long term preservation efforts on our growing collection of microfilm and digital assets. While the Digital Archives and Legacy Washington are both sister divisions under the umbrella of the Secretary of State’s office, the newspapers will continue to be a separately maintained collection of the State Library. The Washington Digital Newspaper collection will only contain the newspaper page images, a contextual essay and minimal descriptive metadata as seen in this example: <https://newspapers.sos.wa.gov/#/newspaper/sn87093220> . WSL will maintain the long term preservation of master TIFFs for disaster recovery. The successful vendor will only need to host the assets needed for web display of the digital collection and restoration.

2. Please describe the current hosting environment. Does the Office have any preference / non-preference when it comes to the hosting of the web portal?

The current hosting environment is limited by our SQL database and the process in which it stores the metadata. The open source viewing software that we were using is not compatible with the SQL server system we currently use for hosting the files.

We do prefer that a company have a data hosting facility in the mainland U.S. and manage collections using U.S. intellectual property standards. Web and content upload support in Pacific Standard/Daylight time zones is also preferred.

3. Is the Office strict about the vendor having "at least 5-10 years of experience in hosting and displaying digital newspaper collections and metadata standards". Is this a mandatory requirement?

Due to the large volume of newspaper content generated by these projects, we have found that companies with 5-10 years of prior experience with newspaper databases and the advance search features needed tend to have a good understanding of the challenges presented. We regularly upload batches of approximately 10,000 pages which are 500Gs each. Consideration will be given to vendors who can show the organized display, ease of use and management of large volumes of newspaper content.