

AMENDMENT 5

QUESTIONS AND ANSWERS

Q33. What is the percentage of your PDF documents that are not digitally created (PDF of scans)?

A33. Currently, 100% of our imaging work come from PDF of scans. We however expect to receive documents electronically (in Word, Excel, Powerpoint or PDF format) specifically for accessibility remediation purposes.

Q34. What is the resolution of your documents (page dimension, DPI)?

A34. Letter or legal size, 200 to 300 dpi based on client's needs.

Q35. Is the diagram posted in the RFP purely a conceptual diagram of how DISC sees this solution being designed or is this the required implementation?

A35. The diagram represents our current operation workflow. However, the two blue boxes describing potential point of implementation for accessibility remediation are purely suggestive.

Q36. What processes in this diagram are currently in place?

A36. All processes except for those in blue boxes.

Q37. Assuming the conceptual diagram is the blueprint for the solution, what components are already in place, and which ones would be expected to be provided by the contractor?

A37. All imaging processes (except for those in blue boxes) are currently in place. Accessibility remediation processes must be added by the contractor.

Q38. Under essential Outcomes, the requirements ask that the solution must “*integrate solid reporting capacity and accessibility validation tools to provide feedback on automated processes.*” – Can you provide examples of what is expected in terms of reporting?

A38. An accessibility validation tool should provide employees with a status of the document processed and indicate success or failure of each action (including, but not limited to, tagging, logical reading order, alternate text and descriptions for images, tables, charts and forms).

Q39. Can you please help us to understand clearly what parts of workflow diagram on Annex 1, are existing today?

A39. Please refer back to answer provided in question 37

Q40. Which parts of existing workflow can be replaced? If none, how we can integrate with existing workflow? What's the system look like? For ex: where is the datacenter; does it have API connector; server type and size; RAM, CPU and OS details?

A40. All the IT systems supporting processes located in the green box reside locally at our Matane installations. Other than preparation and scanning activities, any portion (soft or hard) of this box could be replaced to reach the challenge's objectives. Software and scanning hardware have been detailed in the annex.

Q41. Can we know the details of existing architecture and details of it's components. For ex: What's the existing OCR engine?

A41. A variety of OCR engines, including GD Picture, Abbyy, Nuance and OpenText are used depending on the profile selected for the imaging contract. These choices are not limitative, solutions proposed can make use of any OCR, as long as data territoriality and appropriate handling of protected information is maintained.

Q42. What are the categories will be used in document classification? We want to understand clearly the types of images will be processed.

For ex: only documents like birth certificate or an image of family resting under the tree during the lunch break

A42. Categories are usually document types based on identifiable features (title, barcode, alphanumerical identifier, header page inserted during preparation, etc.) and are used to organise documents for further processing (for example, to exclude specific forms, for ordering purposes or to determine what information to look for and extract on documents depending based on their classification).

Q43. Can we get the historical images with a right classified categories?

A43. This question would be project-specific.

Q44. What existing does QC process involve on each step? What's the definition of good/bad in each QC step?

A44. Current QC process include image quality control and data entry integrity. Image quality control (validating if images have been properly scanned and are not distorted, inverted, backward, removing white pages when

necessary and tagging problematic images for rescan) is done on 100% of images. Data entry integrity control, which validates that data entry (automatic or manual) has been completed properly with no typo or other transcription errors, is applied on a variable scale from 10-100% depending on client's needs. We expect QC of accessibility components to be done on a similar scale depending on complexity of documents and clients' requirements.

Q45. Where the manual classification happen? How to notify a "user" for manual classification?

A45. Manual classification may happen during preparation steps, when documents are getting prepped for scanning. It may also be done later through the process, using a visual interface to manually classify images or PDF documents based on their properties or content.

Q46. What's the policy on using proprietary software as a component in our solution?

For ex: what's the policy about using components from market dominating cloud technology providers like Google, IBM and Amazon (AWS) in order to achieve better scalability (handling high volume of data) and accuracy

A46. Proprietary software can be used as long as (a) it maintains appropriate levels of privacy in regards to transfer and storing of protected information, and (b) data remains in Canada at all times.

Q47. Can you please elaborate more about "extract pertinent and contextualized information from digitized documents"? Can you please give some examples about "pertinent and contextualized information"?

A47. Extraction of document properties or content may provide information about its structure that is pertinent to accessibility remediation. For example, tagging of document sections, reading order, image descriptions and alternate text may potentially be generated from the content and structure of the file itself, thus minimizing manual intervention.

Q48. For the import phase: Can you please discuss and describe the incoming mail, forms, archive documents, fax documents and electronic documents that are the document inputs to your existing system.

Please provide Individual descriptions and indicative samples of each document type. This would be very helpful in our analysis.

A48. Please refer back to answer provided in question 24

Q49. Please discuss if there are any consistency to these incoming documents so that they may be automatically identified as a specific document type for specific down-stream processing. For example, there may be a significant percentage of incoming mail that are completed forms of a specific type and as such could be identified and processed in a specific way. Is this what the automatic classification step does? If so are specific OCR/ICR processing configurations for each of these identified document types in place?

Please provide as many relevant document type samples as you can and the approx. percentage of overall volumes that they represent.

A49. Please refer back to answer provided in question 25

Q50. Are there any consistency of images within or across document types that could possibly automate the addition of alt-text in certain situations? If so, what percentage of image alt-text could be identified and applied automatically today? What is your goal for AI assisted alt-text application within such documents for the future

A50. Please refer back to answer provided in question 27

Q51. What are the average annual and monthly page volumes today and what are the average annual and monthly page volumes expected in the future? Are there any peak volume periods that we should be aware of?

A51. Please refer back to answer provided in question 28

Q52. What are the average required amount of person hours intervention per document is required today after scanning? What is your target expected required future amount of person hours intervention per document when this project is completed?

A52. Please refer back to answer provided in question 29

Q53. Please provide any rules and boundaries associated with tagging an image. Do we have any benchmark reference for evaluating the accuracy of the process?

A53. Attribution of a tag to an image and the determination of an image as decorative or informative can be evaluated as a pass/fail. Alternate texts and description will be evaluated based on their respect of accessibility best practices proposed by the WCAG standard.

Q54. Please provide any rules and boundaries associated with the description of a chart. Do we have any benchmark reference for evaluating the accuracy of the process?

A54. Alternate texts and description will be evaluated based on their respect of accessibility best practices proposed by the WCAG standard.

Q55. Please list all the systems that the provided solution needs to be integrated along with its respective interface availability for each system.

A55. Our current technological infrastructure includes a variety of high-volume scanners such as Fujitsu flatbeds, Exela Intelliscan XDS, Contex Large Format and Widtek Duplex Wide Format scanners. Software-wise, our processes of classification and character recognition are completed primarily on IBM's Datacap as well as Banctec's CenterVision and eFirst solutions.

Q56: In regards to the answer provided to question #20, you indicate *that a validation tool should be provided as part of the quality process assurance.*

Could you please confirm if (a) DISC is going to provide the tool to the supplier; or (b) the supplier should include this tool in their proposed solution?

A56. Supplier should include this tool or use available validation tools to provide post-remediation reporting capacity.

Q57: : In regards to the answer provided to question #31, could you also please indicate the format of the documents once they are scanned (i.e. is it a PDF where each page is an image of a document page, or is it a collection of pages in some images format like jpeg, png, tiff, etc)?

A57. This depends on clients' requirements. In some cases, every page imaged results in an individual PDF file; in others, PDF are assembled from various images depending on preparation and classification rules.

Q58: In regards to the accessibility assessment, could you please confirm if (a) DISC is going to provide the tool to the supplier; or (b) the supplier should include this tool in their proposed solution?

Q58. Please refer back to answer provided in question 59

Q59: In regards to the essential outcome #3 "make possible the conversion of both digitally created and scanned PDF documents into well-structured, properly tagged and fully compliant HTML5 or ePub3 files", our understanding is that tagging is refer to as classification of the documents, is our understanding correct or are your referring to another process?

A59. In the context of imaging, classification refers to a triaging and ordering step not related to accessibility or tagging. This activity may take place during preparation, when documents are getting prepped for scanning, or be completed later through the process, using a visual interface to manually classify electronic images or PDF documents based on their properties or content.