

MODIFICATION 5

QUESTIONS ET RÉPONSES

Q33. Quel pourcentage de vos documents PDF ne sont pas créés numériquement (PDF de documents numérisés)?

R33. Actuellement, toute notre charge de travail liée à l'imagerie est liée à des fichiers PDF obtenus par numérisation. Cependant, le Canada prévoit recevoir des documents par voie électronique (en formats Word, Excel, Powerpoint et PDF), précisément pour en améliorer l'accessibilité.

Q34. Quelle est la résolution de vos documents (dimensions des pages, DPI)?

R34. Format lettre ou légal, 200 à 300 dpi, selon les besoins du client.

Q35. Le diagramme inclus dans la demande de propositions illustre-t-il comment le Centre de solutions en imagerie documentaire conçoit la solution ou représente-t-il sa mise en œuvre exigée?

R35. Le diagramme montre comment se déroulent nos activités à l'heure actuelle. Cependant, les deux cases bleues font état de processus susceptibles d'être mis œuvre pour améliorer l'accessibilité, mais uniquement à titre de suggestion.

Q36. Dans le diagramme, quels processus sont en vigueur?

R36. Tous les processus, sauf ceux présentés dans les cases bleues.

Q37. En supposant que le diagramme conceptuel constitue le plan directeur de la solution, quels éléments sont déjà en place et lesquels doivent être fournis par l'entrepreneur?

R37. Tous les processus d'imagerie (sauf ceux dans les cases bleues) ont déjà été mis en œuvre. Les processus d'amélioration de l'accessibilité doivent être ajoutés par l'entrepreneur.

Q38. Dans la section Résultats essentiels, on précise que la solution doit « comprendre une solide capacité d'établissement de rapports et des outils de validation de l'accessibilité dans le but de fournir une rétroaction sur les processus automatisés. » L'État peut-il donner des exemples de ce qui est attendu en ce qui a trait à la production de rapports?

R38. Un outil de validation de l'accessibilité doit informer l'employé de l'état du document traité et indiquer le résultat – c'est-à-dire la réussite ou l'échec – de chaque intervention, à savoir l'étiquetage du document et

l'établissement de l'ordre logique de sa lecture, ainsi que l'ajout du texte de remplacement et de la description des images, tableaux, graphiques ou formulaires, pour ne nommer que ces actions.

Q39. L'État pourrait-il indiquer plus précisément les éléments du diagramme du flux de travail (Annexe 1) qui sont déjà en place?

R39. Veuillez-vous référer à la réponse fournie à la question 37.

Q40. Quelles portions du diagramme des processus opérationnels en vigueur peuvent être remplacées? S'il n'y a pas, comment pouvons-nous intégrer de nouveaux processus? Quelle est la topologie du système? Par exemple, où se trouve le centre de données? Comprend-il un connecteur API? Quel type de serveur utilisez-vous et quelles en sont les caractéristiques (mémoire vive, unité centrale de traitement, système d'exploitation, etc.)?

R40. Tous les systèmes de la TI qui prennent en charge les processus inclus dans la case verte se trouvent à nos installations de Matane. Outre les activités de préparation et de numérisation, n'importe quel processus (réalisé avec un logiciel ou du matériel) répertorié dans cette case peut être remplacé pour que l'État relève le défi et atteigne ses objectifs. Les logiciels et le matériel de numérisation ont été décrits en détail dans l'annexe.

Q41. Pouvons-nous avoir les détails de l'architecture existante et de ses éléments? Par exemple, quel est le moteur de reconnaissance optique des caractères (ROC)?

R41. Le Canada utilise divers moteurs de ROC, dont GD Picture, Abbyy, Nuance et OpenText, selon le profil sélectionné avec le contrat d'imagerie. Il ne s'agit pas d'une liste exhaustive. Les solutions proposées peuvent faire appel à n'importe quel moteur de ROC, à condition que la territorialité des données et la manipulation appropriée des renseignements protégés soient maintenues.

Q42. Quelles catégories serviront à classer les documents? Nous voulons bien comprendre les types d'images à traiter.

Par exemple, s'agit-il uniquement de documents officiels, comme les certificats de naissance, ou faut-il prévoir une catégorie pour la photographie d'une famille prenant une pause-repas sous un arbre?

R42. En règle générale, les catégories sont définies par les caractéristiques distinctives des documents (titre, code à barres, code d'identification alphanumérique, page titre ajoutée pendant la préparation, etc.). Elles permettent d'organiser les documents en vue de leur traitement ultérieur, que ce soit pour en exclure certains formulaires ou en établir la séquence, ou encore pour faciliter la recherche et l'extraction de l'information qu'ils contiennent.

Q43. Pouvons-nous obtenir les images historiques avec des catégories adéquatement classifiées?

R43. Cette question n'est pertinente que si elle porte sur un projet en particulier.

Q44. Quel processus de contrôle de la qualité (CQ) est appliqué à chaque étape du traitement des images et comment définit-on ce qui est « acceptable » ou « inacceptable »?

R44. Le processus actuel de CQ comprend le contrôle de la qualité des images et de l'intégrité de la saisie des données. Le contrôle de la qualité des images (qui consiste à vérifier si elles ont été numérisées correctement et ne sont pas déformées ou inversées dans un sens ou dans l'autre, à supprimer les pages blanches au besoin et à étiqueter les images problématiques en vue de les numériser de nouveau) est appliqué systématiquement aux images. Le contrôle de l'intégrité de la saisie des données, qui confirme que la saisie des données (automatique ou manuelle) a été correctement effectuée, sans fautes de frappe ou autres erreurs de transcription, est appliqué selon une échelle variable allant de 10 à 100 %, en fonction des besoins des clients. Le Canada prévoit que le CQ des éléments d'accessibilité s'effectuera à une échelle comparable, selon la complexité des documents et les exigences des clients.

Q45. À quel moment a lieu la classification manuelle? Comment informe-t-on un « utilisateur » qu'il doit classer lui-même le document?

R45. La classification manuelle peut être effectuée aux étapes de la préparation des documents à numériser. L'utilisateur peut également s'en charger plus tard dans le processus, en recourant à une interface visuelle qui lui permet de classer lui-même les images ou les documents PDF en fonction de leurs propriétés ou de leur contenu.

Q46. Quelle est la politique relative à l'utilisation d'un logiciel exclusif comme élément de notre solution?

Par exemple, quelle est la politique sur l'utilisation de composantes des fournisseurs de technologies infonuagiques qui dominent le marché, comme Google, IBM et Amazon (AWS), pour améliorer l'évolutivité (la capacité à traiter de grands volumes de données) et la fidélité de la numérisation?

R46. L'utilisation de logiciels exclusifs est acceptable s'ils maintiennent le niveau approprié de protection des renseignements confidentiels transférés ou stockés et si les données restent toujours au Canada.

Q47. Pouvez-vous préciser ce que vous entendez par « extraire des renseignements pertinents et contextuels des documents numérisés »? Pouvez-vous donner des exemples de « renseignements pertinents et contextuels »?

R47. L'extraction des propriétés ou du contenu d'un document peut renseigner sur sa structure et fournir de l'information pertinente pour en améliorer l'accessibilité. Par exemple, l'étiquetage de ses parties, l'ordre dans lequel elles sont lues, la description des images et du texte de remplacement qu'il contient peuvent être générés à partir du contenu et de la structure du fichier lui-même afin d'éliminer le maximum d'interventions manuelles.

Q48. Phase d'importation

L'État pourrait-il décrire le courrier électronique, les formulaires, les documents d'archives, les télécopies et les documents électroniques qui sont traités par votre système actuel?

L'État pourrait-il fournir une description et un échantillon représentatif de chacun de ces types de documents? Cela nous serait très utile pour notre analyse.

R48. Veuillez-vous référer à la réponse fournie à la question 24.

Q49. Les documents à traiter présentent-ils des caractéristiques uniformes permettant de les associer automatiquement à un type particulier et de leur appliquer un traitement précis par la suite? Par exemple, un pourcentage important du courrier reçu pourrait être constitué de formulaires remplis et d'un type particulier qui pourraient donc être identifiés et traités d'une façon particulière. Est-ce là le but de l'étape de la classification automatique? Dans l'affirmative, existe-t-il un traitement de reconnaissance optique ou intelligente des caractères (ROC/RIC) configuré spécifiquement en fonction du type de documents identifié?

L'État pourrait-il nous fournir le maximum possible d'échantillons pertinents des différents types de documents et le pourcentage que chacun représente par rapport au volume global traité?

R49. Veuillez-vous référer à la réponse fournie à la question 25.

Q50. Les images présentent-elles une cohérence dans un même document ou entre documents du même type pour permettre d'automatiser l'insertion de textes de remplacement dans certaines situations? Dans l'affirmative, quel pourcentage de texte de remplacement pourrait être identifié et appliqué automatiquement dès aujourd'hui? Quel est l'objectif de l'État quant à l'application de texte de remplacement au moyen de l'intelligence artificielle dans de tels documents à l'avenir?

R50. Veuillez-vous référer à la réponse fournie à la question 27.

Q51. Quelle quantité de pages l'État prévoit-il traiter en moyenne chaque jour et chaque mois? Y a-t-il des périodes de pointe dont nous devons être au courant?

R51. Veuillez-vous référer à la réponse fournie à la question 28.

Q52. À l'heure actuelle, combien faut-il d'heures-personnes en moyenne pour traiter un document après sa numérisation? Quel est l'objectif de l'État à cet égard une fois le projet terminé?

R52. Veuillez-vous référer à la réponse fournie à la question 29.

Q53. L'État pourrait-il indiquer toute règle et limite associées à l'étiquetage d'une image. Existe-t-il des données de référence pour évaluer l'exactitude du processus?

R53. Le principe de « réussite/échec » peut servir à attribuer une étiquette à une image et à la qualifier de décorative ou d'informatrice. Son texte de remplacement et sa description seront évalués en fonction de leur conformité aux pratiques exemplaires en matière d'accessibilité établies dans la norme WCAG.

Q54. L'État pourrait-il indiquer toute règle et limite associées à la description d'un graphique. Existe-t-il des données de référence pour évaluer l'exactitude du processus?

R54. Le texte de remplacement et la description du graphique seront évalués en fonction de leur conformité aux pratiques exemplaires en matière d'accessibilité établies dans la norme WCAG.

Q55. L'État pourrait-il dresser la liste des systèmes auxquels la solution proposée doit être intégrée, ainsi qu'indiquer si chacun dispose de l'interface nécessaire.

R55. L'infrastructure technologique actuelle du gouvernement du Canada comprend divers scanners à volume de production élevé, comme les appareils à plat Fujitsu, Exela Intelliscan XDS, Contex Large Format et Widtek Duplex Wide Format. Au chapitre des logiciels, les solutions Datacap d'IBM, ainsi que CenterVision et eFirst de Banctec, répondent en grande partie aux besoins du Canada en regard de la classification des documents et de la reconnaissance des caractères.

Q56. En ce qui concerne la réponse fournie à la question #20, vous indiquez: *Un outil de validation devrait être fourni dans le cadre du processus d'assurance de la qualité.*

Pourriez-vous s'il-vous-plait confirmer si (a) CSID fournira l'outil au fournisseur; ou (b) le fournisseur devrait inclure cet outil dans sa solution proposée?

R56. Le fournisseur doit inclure cet outil ou utiliser les outils de validation disponibles pour fournir une capacité de génération de rapports post-correction.

Q57: En ce qui concerne la réponse fournie à la question #31, pourriez-vous également indiquer le format des documents une fois numérisés (c.-à-d. est-ce un PDF où chaque page est une image d'une page de document, ou est-ce une collection de pages dans certains formats d'images comme jpeg, png, tiff, etc.)?

R57. Cela dépend des exigences des clients. Dans certains cas, chaque page imagée entraîne un fichier PDF individuel; dans d'autres, les PDF sont assemblés à partir de diverses images en fonction des règles de préparation et de classification.

Q58: En ce qui concerne l'évaluation de l'accessibilité, pourriez-vous s'il-vous-plait confirmer si (a) CSID fournira l'outil au fournisseur; ou (b) le fournisseur devrait inclure cet outil dans sa solution proposée?

R58. Veuillez-vous référer à la réponse fournie à la question 59.

Q59: En ce qui concerne le résultat essentiel #3 "*permettre la conversion de PDF numérisés et créés électroniquement dans des fichiers HTML5 ou ePub3 bien structurés, balisés et entièrement conformes;*", notre compréhension est que le balisage est appelé classification des documents, notre compréhension est-elle correcte ou faites-vous référence à un autre processus?

R59. Dans le contexte de l'imagerie, la classification fait référence à une étape de tri et de commande non liée à l'accessibilité ou au marquage. Cette activité peut avoir lieu pendant la préparation, lorsque les documents sont préparés pour la numérisation, ou être terminée plus tard dans le processus, à l'aide d'une interface visuelle pour classer manuellement les images électroniques ou les documents PDF en fonction de leurs propriétés ou de leur contenu.