確認手続 Q&A

デジタル化資料の全文テキストデータの視覚障害者等への提供に係る確認手続について、出版社の方々からお寄せいただいた主なご質問にお答えしています。

Q1: 全文テキストデータを提供する資料は、どのように選定されているのでしょうか。

A1: 全文テキストデータは、当館のデジタル化資料の画像データから OCR 処理により作成します。ですので、全文テキストデータを提供する資料は、デジタル化を行った資料を対象としています。当館のデジタル化事業については、「資料デジタル化について」のページをご参照ください。

Q2: 視覚障害者等の方へテキストデータは無料で提供されるのでしょうか。

A2: 当館の視覚障害者等用データ送信サービスに登録された方に対して無料で提供します。

Q3: 全文テキストデータを利用する視覚障害者等の方はどれくらいいるのでしょうか。

A3: 全文テキストデータは、当館の視覚障害者等用データ送信サービスを通じて提供するため、本サービスに登録している個人・機関が対象となります。令和 5 年 3 月末現在で、視覚障害者等個人の方が約 550 名、図書館等が約 240 館になります。

Q4: なぜ提供候補資料リストを確認する作業が必要なのですか。

A4: あらかじめ当館にて、機械的に可能な範囲で除外基準に該当する資料を提供候補から外しますが、当館の調査で把握し切れない場合がありうるため、事前確認手続にご協力を賜りたいと考えております。

Q5: OCR をかけたままの未校正のテキストデータを送信するということですが、認識精度に問題はないのでしょうか。

A5: 当館が提供するデジタル化資料は、昭和前期以前に受け入れた資料が半分近くを占めるという特徴や資料の多様性を踏まえて、当館デジタル化資料に最適化された認識精度と適切な処理速度を確保した OCR 処理プログラムによりテキスト化を行いました。また、令和 4 年度は機械学習技術等の最新技術を取り入れた OCR 処理プログラムの研究開発事業を引き続き行っており、テキストデータの質の改善にも努めています。

令和 5 年 3 月から提供を開始したテキストデータの具体的な認識精度については、NDL ラボ「令和 3 年度デジタル化資料の OCR テキスト化」ページの「5. 性能改善後の文字認識性能（実績値）」をご覧ください。

Q6: 未校正のテキストデータを校正する予定はあるのでしょうか。

A6: 校正済みテキストデータの提供については、当館が別途行っている「学術文献のテキストデータの製作」という枠組みの中で、利用者の方のニーズにお応えしたいと考えています。

Q7: 未校正のテキストデータを提供するということで、出版社側で問い合わせ対応等が必要になる懸念はないのでしょうか。利用者の方への周知はどのように行っていくのでしょうか。

A7: 提供する全文テキストデータの冒頭には注意事項を挿入して、人手による確認を経ていない未校正のテキストデータであることを、利用者の方にご理解いただいた上で利用いただくことを想定しています。膨大な点数を提供することになるため、未校正であることの周知は丁寧に行っていきたいと考えています。また、全文テキストデータは、視覚障害者等用データ送信サービスという、当館が従来から行っている視覚障害者等向けのサービスを通じて提供しますので、お問い合わせはまず当館にお寄せいただくものと想定しております。

Q8: 事前確認手続で申し出なかった場合は、そのまま視覚障害者等用データ送信サービスで提供されることになるのでしょうか。

A8: そのとおりです。なお、提供開始後に、事後確認手続で除外することも可能です。

Q9: 除外基準の中で、「全文テキストデータの内容を全て含む著作物を「同内容の著作物」として除外する」とありますが、再版時に再編集や改変・改訂を行った著作物は「同内容の著作物」に該当するのでしょうか。

A9: 再版時にあらたに内容を追加して再編集を行うなどした場合は、「同内容の著作物」には該当しません。

Q10: 雑誌等で一部のみをデジタル化している場合は除外されますか。

A10: 除外基準は、「同内容の著作物」、すなわち全文テキストデータの内容を全て含む著作物が流通等している場合であるため、一部のみの場合は除外されません。

Q11: 同じ著作者で複数の著作物で構成される出版物（短編集など）について、一部分だけ流通している場合、その出版物全体が提供対象から除外されますか。

A11: 除外されません。除外基準は、「同内容の著作物」、すなわち全文テキストデータの内容を全て含む著作物が流通等している場合であるため、一部のみの場合は除外されません。ただし、文庫版、

作品集・全集、他出版社からの再版等も含め、全ての著作物について同内容の著作物が流通等していることを確認できた場合は除外されます。

Q12: 他社から同内容の書籍が刊行されている場合、元の出版社は除外申出をすることができますか。

A12: 他社から刊行された同内容の書籍が、除外基準に該当して現在流通していれば除外申出をすることができます。

Q13: 提供候補資料リストに自社の名前がない場合、今回はテキストデータが提供されないということでしょうか。

A13: 事前確認手続でご確認いただくのは、原本のデジタル化資料が「国立国会図書館内限定」となっている資料のみとなります。「絶版等（図書館・個人送信資料）」の資料については、「図書館向けデジタル化資料送信サービス（図書館送信）に係る除外手続」において、電子書籍が市場に流通していないことを既にご確認いただいた資料であるため、本事業でご確認いただく「提供候補資料リスト」には含まれておりません。

ですので、提供候補資料リストに名前がない場合でも、絶版等に該当する資料の中に貴社の資料が含まれており、全文テキストデータが提供される可能性はあります。

Q14: 絶版等資料の全文テキストデータ提供に係る事前確認は、また別に作業する必要があるのでしょうか。

A14: 絶版等資料については、「図書館向けデジタル化資料送信サービス（図書館送信）に係る除外手続」において、電子書籍が市場に流通していないことを既にご確認いただいた資料であるため、全文テキストデータの提供に係る事前確認の作業はありません。

Q15: 今回提供する全文テキストデータと国立国会図書館デジタルコレクションで現在提供されている資料との関係がよく分かりません。

A15: 今回提供する全文テキストデータは、令和 2 年 12 月時点で国立国会図書館デジタルコレクションにおいて提供していた資料の画像データから、OCR 処理により作成したものになります。

Q16: 提供候補資料リストのうち、オーディオブックは除外基準①や②の理由では除外申出ができませんが、どの番号で申し出ればよいですか。

A16: オーディオブックについては、除外基準に基づく除外ではなく、送信を留保するという扱いになるので、特に番号は記載せずに「除外申出理由等」の欄に「オーディオブック」と記載してく

ださい。

Q17: 視覚障害者等の方から申出があった場合にテキストデータを提供していることを、自社ホームページ等では明記しておりません。ホームページに明記すれば、今回の提供対象から除外となるでしょうか。

A17: ホームページにてテキストデータをご提供される旨が明記されているということでしたら、除外基準①に該当しますので、ホームページに明記した旨を当館にご連絡いただければ提供対象から除外となります。