

国立国会図書館サーチとの OAI-PMH 連携時に障害となるポイント ～スムーズな連携実現のためにご注意いただきたい点～

(平成 30 年 8 月)

国立国会図書館サーチでは、これまで 100 を越えるデータベースと連携を行ってきました。その経験から、OAI-PMH で連携を開始する際に障害となりうるポイントをご案内します。国立国会図書館サーチとの OAI-PMH でのスムーズな連携実現のために、OAI-PMH リポジトリの実装等にあたって、他の技術資料¹ や OAI-PMH 仕様² と合わせてご参考にしていただければ幸いです。

1 OAI-PMH リポジトリの出力仕様について

1.1 OAI-PMH リポジトリがセッションの維持を要求する

国立国会図書館サーチの OAI-PMH ハーベスタは、`resumptionToken` を使用してリクエストする際にセッションの維持を行いません。そのため、セッションの維持を要求するリポジトリからは `resumptionToken` を使用したメタデータの収集が行えません。

セッションに関わらず、OAI-PMH リクエストを処理できる実装としてください。

1.2 OAI-PMH リポジトリが出力する XML が整形形式でない

国立国会図書館サーチの OAI-PMH ハーベスタは、連携先データベースの OAI-PMH リポジトリに対しリクエストを行い、リポジトリから返戻された XML から `resumptionToken` を特定し、さらに `resumptionToken` を指定したリクエストをリポジトリに送信するという繰り返しにより、メタデータの収集を行います。

この際、リポジトリが返戻する XML が整形形式でないことにより、メタデータの収集に失敗することがあります。具体的な事例は次の通りです。

- XML として許容する文字³ 以外の文字 (制御文字やサロゲートブロック⁴ 等) が出力されている (※国立国会図書館サーチの実装においては文字参照⁵ としての記述や CDATA セクション⁶ としての記述でもエラーとなります)

¹ <http://iss.ndl.go.jp/information/renkei/#5>

² <https://www.nii.ac.jp/irp/archive/translation/oai-pmh2.0/OpenArchivesProtocol.htm>

³ <https://www.w3.org/TR/xml/#charsets>

⁴ U+D800 ~ U+DFFF の領域

⁵ <https://www.w3.org/TR/xml/#dt-charref>

⁶ <https://www.w3.org/TR/xml/#sec-cdata-sect>

- アンパサンド (&) や左山括弧 (<) 等の文字が適切にエスケープ⁷ されていない

リポジトリの返戻は W3C 勧告を満たす整形形式の XML 文書 (Well-Formed XML Document)⁸ である必要があります。

また、国立国会図書館サーチでは私用領域 (Private Use Area)⁹ の文字や Unicode 基本多言語面 (Basic Multilingual Plane; BMP) 外の文字¹⁰ が含まれる XML は正常に扱えません。私用領域・BMP 外の文字が含まれない実装とするようお願いします。

1.3 OAI-PMH リポジトリが出力するメタデータが適切でない

リポジトリが返戻するメタデータが、国立国会図書館サーチのフォーマット仕様 (DC-NDL(RDF)、DC-NDL(Simple)等) に適合していない場合、国立国会図書館サーチへのメタデータの格納や、書誌同定機能等に不具合が発生することがあります。場合によってはリポジトリの出力機能を修正いただく必要があります。リポジトリが返戻するメタデータは国立国会図書館サーチのフォーマット仕様に適合するよう出力していただけるようお願いします。

リポジトリの実装にあたっては DC-NDL(RDF)、DC-NDL(Simple) のフォーマット仕様¹¹ 及び『Web API によるシステム連携ガイドライン』¹² の「2. メタデータ形式」を確実にご確認いただくとともに、提供するスキーマファイル¹³ を用いて妥当性検証をいただければ幸いです。

1.4 OAI-PMH リポジトリがエラー発生時に適切な HTTP ステータスを通知しない

国立国会図書館サーチの OAI-PMH ハーベスタは、連携先データベースの OAI-PMH リポジトリに対しリクエストを行い、正常な返戻が行われなかった場合は処理の中断、あるいはリトライを行います。この判断はリポジトリの通知する HTTP ステータスコードに基づいて行われます。

エラーが発生したにもかかわらずリポジトリが適切なステータスコードを返戻しない (ステータスコード 200 を返戻する) 場合、正常な返戻が行われなかったことを検知できずデータの取り漏れが発生します。

エラー発生時は状況に応じ 500 Internal Server Error、503 Service Unavailable 等の適

⁷ <https://www.w3.org/TR/xml/#syntax>

⁸ <https://www.w3.org/TR/xml/#dt-wellformed>

⁹ U+E000 ~ U+F8FF の領域

¹⁰ U+10000 ~ U+10FFFF の領域

¹¹ <http://iss.ndl.go.jp/information/metadata/#1>

¹² <http://iss.ndl.go.jp/information/renkei/#5>

¹³ 入手をご希望の方はお問い合わせください

切なエラーを出力くださるようお願いします。

1.5 OAI-PMH リポジトリが `resumptionToken` 指定時に仕様外の引数を要求する

OAI-PMH の仕様上、`resumptionToken` は排他 (exclusive) の引数として定義されている¹⁴ため、OAI-PMH ハーベスタが `resumptionToken` を指定してリクエストを送信する場合は、他の引数 (`from`, `until`, `set`, `metadataPrefix`) は付加しないのが正しい仕様となります。

そのため、リポジトリが他の引数がない場合に正常にリクエストが処理されない実装となっている場合は、リポジトリ実装の修正を実施いただく等の対応が必要となり、連携開始までに時間を要することとなります。

`resumptionToken` のみを引数に指定したリクエストについて、正常に処理できるリポジトリ実装とするようお願いします。

1.6 OAI-PMH リポジトリが `resumptionToken` 指定リクエストに対し冪等性がない

OAI-PMH の仕様¹⁵では、同一の `resumptionToken` を指定してリクエストを行った場合に、同じレコードの集合を返す必要があります (リポジトリのレコードに変更が生じた場合を除く)。同一の `resumptionToken` に対して返されるレコードが試行ごとに変わる実装の場合、失敗したリクエストの再試行が行えず、メタデータの取得に失敗する原因となります。

リポジトリのレコードに変更がない限り、同一の `resumptionToken` を指定したリクエストに対し、同一のレコードの集合を返すようお願いします。なお、`ListRecords` リクエストに対する返戻中の `resumptionToken` 要素に `completeListSize` 属性、`cursor` 属性が含まれていると、`ListRecords` リクエストの処理に問題が発生した際にハンドリングが容易となります。可能な範囲でご対応いただければ幸いです。

1.7 OAI-PMH リポジトリの返戻性能に問題がある

国立国会図書館サーチとの連携にあたっては、連携先データベースの OAI-PMH リポジトリに対して国立国会図書館サーチの OAI-PMH ハーベスタからリクエストを送信し、OAI-PMH リポジトリがメタデータを返戻するというやりとりが行われます。

この際、リポジトリの返戻性能が低くリクエストの処理に長い時間がかかることが、連携

¹⁴ <https://www.nii.ac.jp/irp/archive/translation/oai-pmh2.0/OpenArchivesProtocol.htm#ListRecords>

¹⁵ <https://www.nii.ac.jp/irp/archive/translation/oai-pmh2.0/OpenArchivesProtocol.htm#Idempotency>

に支障となっている場合がございます。

実際に必要となる返戻性能は、1回の収集で出力される更新レコード数等によっても変わってきますが、1回のリクエストに対し返戻まで2分以上かかる場合は、国立国会図書館サーチとの連携上支障があります。(1回の収集で出力される更新レコード数が多い場合は、より高い返戻性能が要求されます。)

また、返戻までの時間が出力対象のレコード数に応じて増加するようなケースについては、とくに余裕を持った返戻性能を担保いただきますようお願いいたします。(連携開始時の想定を超える数の更新レコードが出力された場合、タイムアウト等でメタデータの取得に失敗する可能性があります)。

国立国会図書館の公開する『OAI-PMHの要点』¹⁶の「2.5 正確性と速度」及び「3.5 出力するXMLデータの生成タイミング」を参考に、想定される出力件数に応じ、連携上問題とならない範囲の時間でリクエストに返戻できる性能を満たすようにリポジトリを実装いただけるようお願いいたします。

また、特定の `datestamp` に多量のレコードが出力される場合には、出力レコードの `datestamp` を平準化することによって、1回の収集における出力件数を調整いただけるよう依頼することがございます。

1.8 OAI-PMH リポジトリが削除レコードをサポートしていない

OAI-PMH リポジトリが削除レコードをサポートしていない場合、定期的なレコードの全件入替作業が必要となり、OAI-PMHによる連携の重要なメリットである当館・連携機関双方の運用負荷の軽減が実現できないこととなります。また、システムへの負荷が高くなるため運用設計等について当館と連携機関との間で調整の必要があり、連携開始までの時間を要します。

原則として OAI-PMH リポジトリで削除レコードの出力(`persistent` ないし `transient`)をサポートいただくようお願いいたします。

1.9 OAI-PMH リポジトリが日単位の選択的ハーベストをサポートしていない

OAI-PMH リポジトリが選択的ハーベスト¹⁷をサポートしていない場合、収集のたびにデータを全件取得することとなり、特にデータ件数が多い場合は当館・連携機関双方にとって運用上の負荷となります。なお、OAI-PMHでの仕様では選択的ハーベストをサポートするリポジトリは日単位での選択的ハーベスティングをサポートするものとされています。

¹⁶ <http://iss.ndl.go.jp/information/renkei/#5>

¹⁷ <https://www.nii.ac.jp/irp/archive/translation/oai-pmh2.0/OpenArchivesProtocol.htm#SelectiveHarvesting>

原則として日単位の選択的ハーベストをサポートいただくようお願いします。

2 その他

2.1 全件データの媒体提供ができない場合

連携開始時は大量のデータを国立国会図書館サーチに投入する必要があるため、全件データの媒体提供をお願いしております。

連携先データベースの仕様で出力ができない等の事情がある場合は、当館にて OAI-PMH でのデータ収集を検討いたしますが、連携先のデータ件数や返戻性能によっては、当館による全件データ収集が不可能あるいは非常に長期間を要する場合がございます（総合目録ネットワークのように件数が多い場合は OAI-PMH での全件取得が困難な場合が多く見られます）。

原則として、全件データの媒体提供が可能となるようにご準備をお願いします。

2.2 遷移先リンクが永続的でない

国立国会図書館サーチからの遷移先リンクの URL が永続的でない場合、システムリニューアル等で遷移先が変更になった際に、連携済みメタデータの全件更新が必要となります。

遷移先リンクとして使用する URL は永続的なものとするをおすすめします。