

# 大学の図書館

第43巻第4号 (No.605)

2024 4



## 目次

新しいアプリや機能 ..... 楫 幸子 ... 47

特集：OpenRefine

    特集：『Open Refine』編集にあたって ..... 野間口真裕 ... 48

    OpenRefine 活用事例 WebAPI で取得した情報を加工する ..... 伊東 洋輔 ... 53

    OpenRefine を図書館業務で使ってみて ..... 松野 高德 ... 54

    複数著者のリポジトリ登録に便利な OpenRefine ..... 楫 幸子 ... 57

    データスキル勉強会へのお誘い ..... 田辺 浩介 ... 60

大学図書館研究会関西3地域グループ合同例会  
「ネットで探せない『書誌の書誌』の書籍化の裏話」開催報告 ..... 山上 朋宏 ... 61

議事要録 ..... 62

組織通信 ..... 62

### 新しいアプリや機能

楫 幸子

今号の特集でOpenRefineの活用例を紹介しました。とても便利なアプリです。

これまでも作業を効率化してくれる画期的なアプリや機能が色々ありました。しかし次から次へと出現する新技術を使いこなしているかという、否です。

業務に必須と思われるExcelに関して、せいぜいマクロの記録を活用する程度で、その登録したマクロですら、日が経つにつれて前提条件や中身の作業を忘れていて、エラーが出たときに対応できなかつたりします。関数が多用された表は、仕事を引き継いだときに問題になります。

今回OpenRefineという新しいアプリを使ってみたのですが、複数の著者をいとも簡単に各カラム(列)に分割してくれました。リポジトリの一括登録において大いに役立つアプリです。ただ、何年か後のことを考えたとき、有料化や更新停止などもありこのアプリが変わらず

に使用できるとは限りません。その場合、ここをクリックしてこれを選択するという作業手順だけのマニュアルでは不十分です。作業の前条件や目的などもきちんと文章化する必要があります。それらの記録があれば、次の担当者はその時の新しいアプリで対応するかもしれませんし、もしくはChatGPTに聞いてみるかもしれません。

コンピュータ関連の技術やサービスは変化が激しく、せつかくの自動化や効率化も他の人へ引き継げるのか不安を覚えることがあります。しかし、これまで苦労していた作業が驚くほど一瞬で完了することもあります。これからは手に負えないと思いつつ、色々試していけたらと思います。

最後に、学術基盤整備研究グループではこのOpenRefineなどデータスキル勉強会をオンラインにておこなっています。ぜひみなさまのご参加をお待ちしています。

(かじ・さちこ／安田女子大学図書館)

---

## 特集：OpenRefine

---

本特集では当グループが行っているデータスキル勉強会で教材としておりました「Library Carpentry」ならびに第54回全国大会第6分科会「データクレンジングのはじめ方」で実習いたしましたソフトウェア「OpenRefine」の活用事例を紹介いただきました。

図書館が扱う多様なメタデータ整備の一助となりますと幸いです。

(特集企画担当：学術基盤整備研究グループ)

---

### 特集：『Open Refine』編集にあたって

野間口 真裕

Open Refineは、データのクリーニングや変換、分析に優れたツールです。海外の図書館コミュニティではしばしば使用されています。しかし、日本では機能や使い方はまだ十分に知られていません。この特集号では、Open Refineの基本的な操作方法や勉強方法やTips、実践例を紹介します。これからメタデータを扱う図書館職員にとって、Open Refineは強力なツールとなるでしょう。ぜひ、この号から、Open Refineの魅力に触れてみてください。

Open RefineはFreebaseGridworksという名前でMetawebという会社が開発していました。2010年にGoogle, Inc.がMetawebを買収し、FreeGridworksをGoogle Refineと改名しました。しかし、2012年にGoogleはGoogle Refineのサポートを終了し、オープンソースコミュニティに引き継がれました。それ以降、Open Refineという名前となっています。

URL：<https://googleblog.blogspot.com/2010/07/deeper-understanding-with-metaweb.html>

OpenRefineはBSDライセンスによるフ

リーのツールです。最新版はWebに公開されており、OpenRefineのページからダウンロードできます。

Googleの検索結果などからOpenRefineのページにいきましょう。英語ですが、ダウンロードやマニュアルはこちらから情報を得ることができます。

URL：<https://openrefine.org/>

トップページの水色のボタン（Download）からDownload OpenRefineのページに移動します。その後、水色のボタン（Download OpenRefine 3.7.9 for Windows）をクリックしてダウンロードします。すると圧縮されたzipファイルがダウンロードされますので、ダウンロードされたファイル上で右クリックしてすべて展開を選択します。指定された場所に解凍して展開されます。openrefine.exeをクリックしたらOpenRefineがブラウザで開きます。

OpenRefineの主な特徴はファセットを使用した大規模なデータセットのフィルター処理、ヒューリスティック（発見的手法）によるクラスタリング、図書館や学術機関が整備したデータベースとの照合作業（名寄せ）、バージョンの操作履歴を使用した元に戻してのやり直し、自分のマシン上でのクリーニング作業があげられます。フィルター処理は



といった画面が現れます。これはJavaがOpen Refineのアプリケーションを起動していますので、画面の遷移を見ながら少し待ってください。すると既定のブラウザが開いて、ダイヤモンドのアイコンのOpenRefineの画面が開きます。URL欄は127.0.0.1:3333になります。

OpenRefineのロゴの下にメニューがCreate project、Open project、Import Projectと並んで、Language settingsがあります。こちらをクリックして、Select preferred languageから日本語を選ぶと、Change languageのボタンで表示言語を変更できます。

データ形式としてはTSV, CSV, \*SV, Excel (.xls and .xlsx), JSON, XML, RDF as XML, and Google Data documentsをサポートしています。テキスト形式の大規模データであれば、パソコンの能力次第で大量に読み込むことができます。Create project(新規プロジェクト)からはじめます。

データをアップロードするとOpenRefineがデータ解析します。規則的なデータであれば、ほぼスムーズに変換されますが、特殊文字など不規則なデータが含まれる場合はデータの前処理が必要かもしれませんし、先頭行や空白文字を削除するだけで、規則的なデータになるかもしれません。

真ん中に表示されている白と灰色のストライプの表が解析結果のプレビュー画面ですのうまく表示されたら、Project name(名前)を記入してCreate project(プロジェクトを作成)します。

ここからのデータ処理は自分の用意したデータ次第ですので、それは事例紹介の他稿をみていただくとして、当グループが確認している勉強法をお伝えします。

大学図書館研究会第54回全国大会第6分科会にて「データクレンジングのはじめ方」という題名で前田朗様に講師をしていただき、OpenRefineを実習しました。当日の発表資料2.実習OpenRefineの使い方\_縦長.pdfに詳しく掲載いただいています。

URL : <https://sites.google.com/site/scienceinfra/全国大会分科会/第54回全国大会第6分科会?authuser=0>

また、国立情報学研究所大学図書館員のためのIT総合研修にて「2023年度 テーマ：WebAPIを使ったデータの入手とその整備」が開催されました。図書館Web APIのためのOpenRefine活用法(2)に詳しく、事例が紹介されています。

URL : <https://contents.nii.ac.jp/hrd/it/2023/result>

図書館および情報関連のコミュニティ内でソフトウェアとデータのスキルを構築することに重点を置いた学習教材Library CarpentryのOpenRefineのレッスンで説明文章を確認することができます。英語ですが、左メニューのSee all in one pageで1ページにして、ブラウザの機能で日本語に翻訳していただいたら雰囲気をつかむことができますので参考にご利用ください。

URL : <https://librarycarpentry.org/lc-openrefine/>

最後に会員の松野氏にOpenRefineの機能一覧を作成いただきましたので、CheetSheetとして、ご紹介いたします。また、このあとのページにて、伊東氏、松野氏、楫氏に事例を紹介いただきました。ぜひご自身の職場でもお試しく下さい。データのクリーニング、解析に役立ちますと幸いです。

(のまぐち・まさひろ/京都大学附属図書館)

## 付録：OpenRefine 機能一覧

<全体>
Transform...：変換
Edit all colmns：全カラム編集
Trim leading and trailing whitespace...：先頭と末尾の空白を削除...
Collapse consecutive whitespace...：連続した空白を縮める...
Unescape HTML entities...：HTMLをアンエスケープ...
Replace smart quotes with ASCII...：スマート引用符置換...
To titlecase...：語頭だけ大文字に...
To uppercase...：大文字に...
To lowercase...：小文字に...
To number...：数値に...
To date...：日付に...
To text...：文字列に...
To null...：セルをnullに...
To empty string...：セルを空白に...
Facet：ファセット
Facet by star：星ファセット
Facet by flag：旗ファセット
Facet by blank(null or empty string)：空白ファセット(null/空)
Blank values per column：カラムの空白値
Blank records per column：カラムの空白レコード
Non-Blank values per column：カラムの非空白値
Nonblank records per column：カラムの非空白レコード
Edit rows：行を編集
Star rows：星を付ける
Unstar rows：星を外す
Flag rows：旗を立てる
unflag rows：旗を外す
Remove matching rows：マッチした行を削除
Edit columns：カラムを編集
Reorder/remove columns...：カラムの並べ替え・削除...
Full down：下方向の空白を埋める
Blank down：下方向に空白にする
View：ビュー
collapse all columns：すべてのカラムをたたむ
Expand all columns：すべてのカラムを開く
Show/Hide null alues in fells：null値を表示/非表示
<個別項目>
facet：ファセット
Text facet：文字列ファセット
Numeric facet：数値ファセット
Timeline facet：タイムラインファセット
Scatterplot facet...：散布図ファセット...
Custom text facet...：カスタム文字列ファセット...
Custom numeric facet...：カスタム数値ファセット...
Customized facets：カスタムファセット
Word facet：単語ファセット

Duplicates facet : 重複ファセット
Numeric log facet : 対数ファセット
1-bounded numeric log facet : 対数ファセット (1以上を強制)
Text length facet : 文字数ファセット
Log of text length facet : 文字数対数ファセット
Unicode char-code facet : 文字コードファセット
Facet by error : エラーファセット
Facet by null : nullファセット
Facet by empty string : 空の文字列ファセット
Facet by blank(null or empty string) : 空白ファセット (null/空)
Text filter : 文字列フィルタ
Edit cells : セル編集
Transofrm... : 変換...
Common Trsansforms : よく使う変換
Trim leading and trailing whitespace
Collapse consecutive whitespace
Unescspe HTML entitles
Replace smart quotes with ASCII : スマート引用符置換
To titlecase : 語頭だけ大文字に
To uppercase : 大文字に
To lowercase : 小文字に
To number : 数値に
To date : 日付に
To Text : 文字列に
To null : セルをnullに
To empty string : セルを空白に
Full down : 下方向の空白を埋める
Blankdown : 下方向に空白にする
Split multi-valued cells... : 多値のセルを分割...
Join multi-valued cells... : 多値のセルを結合...
Cluster and Edit... : クラスタと編集...
Replace : 置換
Edit column : カラム編集
Split into servarl columns... : 複数のカラムに分割...
Join columns... : カラムの結合
Add column based on this column... : このカラムに基づいてカラムを追加...
Add column by fetching URLs... : URLでカラムを追加...
Add columns from reconciled values... : 照合値からカラムを作る...
Rename this column... : カラム名を変更する...
Remove this column : このカラムを取り除く
Move column to beginning : 先頭にカラムを移動
Move column to end : 末尾にカラムを移動
Move column left : 左にカラムを移動
Move column right : 右にカラムを移動
Transpose : 行列転置
Transpose cells across columns into rows... : 列を行に転置 (縦持ち化) ...
Transpose cells in rows into columns... : 行を列に転置 (等間隔で横持ち化) ...
Columnize by key/value columns... : 行を列に転置 (値に応じた横持ち化) ...

Sort : ソート
View : ビュー
Collapse the column : カラムをたたむ
Collapse all other columns : これ以外のカラムをたたむ
Collapse all columns to left : 左側をすべてたたむ
Collapse all columns to right : 右側をすべてたたむ
Expand all columns : すべてのカラムを開く
Expand all columns to the left : カラムを左側に展開
Expand all columns to the right : カラムを右側に展開
Reconcile : 照合 (名寄せ)
Start reconciling : 照合開始 (reconcile)
Facets : ファセット
by Judgment : 照合判断 (judgement)
Judgment action type : 判断アクション
Judgment action timestamp : 判断履歴
Best candidate's score : 最良候補のスコア
Best candidate's type match : 最良候補の型のマッチ
Best candidate's name match : 最良候補の名前のマッチ
Best candidate's name edit distance : 最良候補の名前の編集距離
Best candidate's name word similarity : 最良候補の名前の単語類似比率
Best candidate's types : 最良候補型
Match each cell to its best candidate : 最良候補と各セルをマッチさせる
Create a new item for each cell... : 各セルから新しい項目を作成...
Create one new item for similar cells... : 同内容のセルの新しい項目を作成...
Match all filtered cells to... : 選択したセルとマッチさせる対象は...
Discard reconciliation judgments : 照合判断を捨てる
Clear reconciliation data : 照合データをクリア
Copy reconciliation data... : 照合データをコピー ...
Use Values as identifiers... : 値を識別子として使用...
Add Entity identifiers column... : 項目識別カラムを追加...

## OpenRefine 活用事例 WebAPI で取得した情報を加工する

伊東 洋輔

### OpenRefine を使い始めたきっかけ

筆者は機関リポジトリの運用という、デジタルと大いに関わる業務を主担当としていることもあり、DX化・効率化に比較的近い環境で働いている。現在、特に作業量が多い業務として、科研成果報告書に記載された成果物の集計がある。これの手順は、報告書のPDFファイル上のテキストデータの必要箇所をExcelに転記するというもので、要する

にマウスの操作を主としたコピー作業の繰り返しである。この非効率な状況を打開したのがOpenRefineである。

### OpenRefine の導入

科研報告書に記載された内容は、科学研究費助成事業費データベースKAKENでデジタル化されており、WebAPIを利用することでデータを取得可能である。しかし、WebAPIによって得られるデータの形式はxmlやjsonといった、専門的なスキルや環境が無いと利用が難しいものである。似た用途で使われるものとしてはcsvがあるが、これはxmlや

jsonに比べると柔軟性に乏しい一方、Excelにインポートするだけで容易に人間が理解しやすい形式に整えられるという利点がある。OpenRefineは、「人間が理解しやすい形式」にデータを加工することを得意とするツールで、csvのみならずxmlやjsonまで簡単な操作で表形式に加工することができる。

なお、WebAPIで得られる生データは対象物に関する情報の全てが出力されるため、例えばKAKENの場合、特定の課題番号の情報を出力すると報告書に記載されている内容を全て得られる。この場合、業務として必要なのは成果「論文」であっても、学会発表等、媒体の異なる成果や、研究報告そのものや研究者のプロフィールといった情報も一律に取得される。OpenRefineではこれにもある程度対応可能で、取り込んだデータの特定の階層のみを表として整形することが可能となっている。これによって、取得した生データから論文の情報が記載されている部分のみを表にすることも出来る。

### 得られた効果

OpenRefineの導入によって、当初期待された通り報告書に記載された情報をExcelに転記する作業に要する時間は大幅に短縮された。これまでのPDFからコピペする作業の場合、報告者が記載した成果の件数によって一報告書あたりにかかる時間が大きく異なるため作業時間の見通しを立てることが不可能だったが、一連の操作で報告書に記載された全ての情報を取り込むことができ、作業時間を報告書の数から予測することができるようになる。

このように、「報告書の内容をExcelで表にする」ことは大いに効率化されたものの、内容の精査については依然として課題が残る。何故なら、報告書に記載された内容が完全でなかったり誤っていることもあるため、それらの確認はこれまで通り人力で行なう必

要がある。

### おわりに

筆者の業務においてOpenRefineを導入して1年目ではあるが、業務の時短について大きな効果を得られている。把握しきれていない機能や、KAKEN以外のデータベースとの連携（CrossRefやCiNii Researchから正確な論文情報を取得するなど）による効率化の余地が残されている。

また、OpenRefineそのものは直接業務効率化には繋がらず、本学ではWebAPIによるKAKENからの情報取得という前提があるように、xmlやjsonの、活用できれば便利だが利用方法が分からないというデータを見つけることが先になる。また、csvはExcelで直接加工しやすいデータ形式ではあるが、稀にExcelへインポートするとレイアウトが崩れてしまう記述になっているものもある。そのようなものもOpenRefineに取り込むと、崩れず展開できることがあるので、覚えていて損はないツールだと思われる。

(いとう・ようすけ／信州大学附属図書館)

## OpenRefine を図書館業務で使ってみて

松野 高德

### 1. はじめに

利用している学校図書館システムにある書誌データは、CSV形式のデータで出力ができることから、出力したCSVファイルを開いてOpenRefineに取り込むことができた。図書館システムに格納されている書誌データの件数は、5万件弱で、その内の約1万件は除籍した書誌データである。

今回、除籍していない約4万件の書誌データについて、OpenRefineを利用しどのよう

なものであるかを把握してみた（全国学校図書館協議会の全国学校図書館調査2023年（悉皆調査でなくサンプリング調査）では、高校図書館の平均蔵書冊数は26,534冊とのことで、OpenRefineで操作するには手頃のデータ件数なのかもしれない。）。結論は、オリジナル入力したデータにかなり乱雑なデータ（後述）があった。

利用している図書館システムには、OPAC機能がないので、カーリル学校図書館支援プログラムを利用して簡易OPACを提供している。生徒たちが目にするようになる書誌データに乱雑なデータがあっては不都合なので、データクレンジングで整った書誌データに整備できたらと、作業を進めてみた。

## 2. 書誌データの傾向把握

OpenRefineでデータを取込み、その後「プロジェクトを作成」を選択すると、データクレンジングを始める準備が整う。この段階で、表示がレコードになっているので行モードに変更することをお勧めする。この表示のレコードモードから行モードに変更することに気づくまで、取込んだデータ件数と元のデータの件数が一致しないので、正しく取り込まれていないと考え、何度もデータの取り込みを繰り返し、無駄な時間を費やした。

取込んだデータは、「文字列ファセット」を使うことで、データのパターンやトレンドが把握できる。OpenRefineのダイヤモンドのロゴはそこから来ているとのこと（ここで使っている「ファセット」の語義をわかりやすく説明するのは至難かもしれない。）。

現在の目録作業は、TRC MARCを利用して、何時から利用しているかを調べてみた。取込んだレコード項目にTRCMARCNOと受入日があったので、この2項目を使い、TRCMARCNO列に空白ファセットを適用しfalse（TRC MARCの番号を有するデータを抽出）を選択、その後で受入日の列に文字列

ファセットを適用する。その結果、受入日ごとのTRCMARCNOを有するレコード件数が左側に受入日順に整列されて表示される。表示結果を見ると、TRC MARCの番号を有する最初の図書は、2016年6月2日であった。（現在の目録作成は、TRCのTOOL-iSを使用している。）」

文字列ファセットでTRC MARCの番号を有するレコードを選択し、文字列ファセットを出版社、出版年、判型、頁数の各項目を適用すると、データに乱雑さはなく綺麗なデータが表示され、データクレンジングの必要性を認めなかった。他方、文字列ファセットでTRC MARCの番号を有しないレコードを選択し、文字列ファセットを出版社、出版年、判型、頁数の各項目を適用すると、次のとおり乱雑なデータがあった。

- ① 出版社の文字列ファセット 英数字の半角と全角の混在、漢字の誤変換、出版社名が途中で省略されて欠損等。
- ② 出版年の文字列ファセット 4桁でなく2桁や3桁の数字、叢書の巻号にあたる数字の誤入力等。
- ③ 判型の文字列ファセット センチ等の単位記号が無い、数字の半角と全角の混在、本の大きさなのに「p」を有するデータ等。
- ④ 頁数の文字列ファセット 頁数の単位記号「p」が無い、本の頁数なのに「cm」を有するデータ等。※頁数のデータについて、図書にローマ数字（例えば、「iv」とか「xxv」）による頁の記載をアラビア数字で表記にしていることに気がついた（TRC MARCも同様）。

元データを変更せずに文字列ファセットを適用するだけでなく、文字列を分割した結果に基づき文字列ファセットを適用することも可能である。例えば、叢書名のデータについて、カラム→複数カラムに分割を選択し、分割をする区切り文字を全角空白として、「分

割後のカラムを取り除く」のチェックを外して処理すると、叢書名と巻号を分離する全角空白の文字のないデータ（叢書名と巻号に区切り記号が無く結合した状態）があった。

請求記号は、ローカル情報なので、オリジナル入力データも TRC MARC 利用データも同様で、次のように乱雑なデータがあった。

- ① 分類記号 分類番号を4桁目の基準に substring して文字列を分割、4桁目のデータについて文字列ファセットを適用すると、「.」以外の文字があり、その内容は0類での頭の0の欠損（過去に EXCEL 保存したファイルをアップロードした可能性がある。）や「.」の位置誤り、登録番号の誤入力があった。
- ② 図書記号 半角カナと全角カナの混在、誤入力によると思われるカナ以外の文字の誤入力があった。

### 3. OpenRefine を操作して気づいたこと

CiNii で検索しても OpenRefine に関する日本語図書はヒットしない。学術基盤整備研究グループで2023年夏頃から始めた読書会 (Introduction to OpenRefine <https://librarycarpentry.org/lc-open-refine/01-introduction.html>) や昨年9月の全国大会第6分科会で得た情報で作業をしている。わからないときは、OpenRefine user manual (<https://openrefine.org/docs>) を拾い読みしながら作業を進めている。このマニュアルをきちんと読めば書かれているのかもしれないが、操作をしていて気がついたことは、次のとおり。

- ・ OpenRefine は、shiftjis の文字コードのファイルも読み込んでくれる。他方、OpenRefine のデータをエクスポートした CSV ファイルの文字コードは UTF8。利用している図書館システムのデータの取り込みは、shiftjis 対応で、文字コードを UTF8 から shiftjis に変更が必要になる。
- ・ セルの幅を広げるには、「カラム編集→カ

ラム名を変更する」を選択し、カラム名に適当な文字を追加することで追加した文字数分の幅が広がる。ただし、空白だけ入れてもカラム名の幅は変わらなかった。

- ・ 文字列データを数字に変換するには toNumber 関数を利用する。データを数字に変換すると、表示データの色が緑色に変わった。

### 4. おわりに

利用している図書館システムには、ローカルで作成した書誌データ (CSV 形式) をアップロードして取り込む機能がある。この機能を使い OpenRefine で修正した書誌データをアップロードし、図書館システムのデータとして修正を反映できる。

修正作業は、「edit」で修正内容を「同じ内容の全セルに適用」とすることができるので、同じ内容の修正であれば、図書館システムで1件ごとに修正するのと比べて、レコードを抽出する手間が省けるので修正作業を効率よく進めることができる。レコードを修正する際は、修正する前か後に OpenRefine 左端の全ての箇所列にある星マークを黄色にし、後で星ファセットの星マークの黄色を抽出してエクスポートすれば修正結果のみのファイルが作成できる。

ただ、乱雑なデータの多くは、オリジナル入力したものにあり、各レコード項目に文字列ファセットを適用した結果を眺めていると、著者名項目データに「著」や「編」の記載がない、叢書名項目データにヨミがないものが大量にあり、図書に記載されているとおりには叢書名が入力されていない、ISBN 項目で同一の ISBN で書名が異なるという ISBN の誤入力がある等の不具合が目についてしまう。これらの書誌情報の修正は、該当図書による確認が必要となるので OpenRefine の利用による修正は困難と思われる。

今回の作業では、規則性のある乱雑なデー

タであればOpenRefineを利用すれば効率的に修正作業が可能であること、書誌データの乱雑さの程度をレコード項目ごとに把握できたこと、オリジナル入力した書誌データの整備をすれば揃ったデータとなることを確認することができた。

(まつの・たかのり／多治見西高等学校)

## 複数著者のリポジトリ登録に 便利な OpenRefine

楯 幸子

### 1. はじめに

学内学会誌をリポジトリに載せたいという話が教員からあり、どのようなデータを用意すればいいのかという問い合わせがあった。すでにリポジトリには学位論文と大学紀要を掲載していたが、その都度必要なデータを集めていた。しかし今後はさまざまな紀要を掲載する可能性もある。できるだけ一括登録がしやすい形式を検討した。

依頼するデータに必須としたのが、著者名やキーワードなど複数入力可能な項目に、あらかじめ区切り記号を入力してもらうことだ。最初はその記号によりExcelの区切り位置機能を使う予定だった。しかしOpenRefineを使えばより簡単に作業ができそうだと思う取り入れてみた。

これが最適な方法とは言えないが、提出されたリストの共著者が何人いるのか不明で、どこまで列を増やせばいいのか試行錯誤したことがある人はぜひ情報を共有したい。

### 2. 依頼する紀要のメタデータ

本学のリポジトリはJAIRO Cloudを利用している。提出された論文リストは、そのままコピーして一括登録用のメタデータに流用したいので、特に重視したのが以下の3点だ。

- Excelファイルで、全てのセルの表示形式が文字列になっている。
- 複数登録できる著者名やキーワードは区切り記号パイプ (|) を使用する。
- 著者名は姓、カンマ (,)、半角空白、名前の形式で入力する。

見本として用意したExcelファイルは、初めから表示形式を文字列とした。入力の注意点をシートの1行目に記入し、例えば区切り記号にパイプ (|) を使い、前後に空白を入れないことなどを注記した。著者の書き方もミドルネームを含めて説明した。言語設定も可能なところは入力してもらったが、日本語 (ja) と英語 (en) 以外は言語名をそのまま書いてもらうようにした。

2行目には見出しを入れた。この見出しは一括登録で使われる見出し (内部キーやラベル) ではなく、一般に使われる語句にした。3行目と4行目には入力例を入れて、日付なども最初から「2024-04-01」の形式で入力してもらった。

Excelシートは横に長くて全体像をつかめないため、入力する項目を説明した表もセットで提供した。

細かい入力規則を提示したので、本当にその通りに作業してもらえるか不安だったが、提出されたリストはほぼ問題がなかった。著者名もすべて規則通りに入力されていて大変有難かった。ただ、ISSNの入力欄は、PISSNとEISSNという見出しにしていたから空白だった。冊子と電子ジャーナルのISSNという追加の説明が必要だろう。

### 3. 一括登録用のメタデータ

JAIRO Cloudを利用する上で、WEKO2からWEKO3にシステムが変わり一番とまどったのは、複数の著者名をひとつのセルにまとめて入力できないことだった。共著者がいる場合、人数分の列を追加しなければならない。しかも著者名に対応する言語の列を追加し

て、0から始まる見出しの数字も的確に増やす必要があった。最初の頃ひたすら手で列を挿入していたためか、よくサーバー内部エラーになっていた。

現在の本学における紀要論文を一括登録する手順は以下ようになる。OpenRefineでのセルの分割と、見出しの連番設定については後で詳しく説明する。

なお、本学では紀要論文を掲載順に登録したいため、仮登録をおこなっている。記述は省いたが、一度必要最低限のメタデータで仮登録をおこない掲載順を整えてから以下の作業を実施している。

- ① 紀要論文のメタデータファイルをエクスポートしてExcelで開く。
- ② 提出されたデータからコピーする等、一括登録用にメタデータを整えExcel形式で保存する。その際、キーワードや著者名など複数登録できる項目は、区切り記号付きでそのままひとつのセルに入力しておく。
- ③ OpenRefineを使いセルを分割する。
- ④ Excel形式で出力して、増加した列の見出しを作成する。
- ⑤ タブ区切りのメタデータファイルに変換して、一括登録をおこなう。
- ⑥ 登録したアイテムをエクスポートして、キーワードや著者名の言語を追加する。
- ⑦ 再度メタデータの一括登録をおこなう。

エクスポートしてデータを追記後インポートする作業が何度もあるが、実際には本文ファイルの登録やDOIの自動採番などの作業と組み合わせておこなっている。

#### 4. OpenRefineにてセルを分割

複数の著者をセルごとに分割する作業は、OpenRefineをインストールするだけで始められる。アプリの表記は日本語へ変えることができ、オフラインで利用できる。元のデータに直接変更を加えずに作業可能だ。上記3-

②で保存した一括登録用のExcelファイルを使用する。

#### 4.1 データ解析

OpenRefineを開き、新規プロジェクトにてExcelファイルを選択すると、解析オプションを設定する画面が表示される。後の見出しの操作のため、内部キーとラベルの行をカラム名として扱うとやりやすい。以下のように設定する。

- ・ データの解析形式：Excelファイル
- ・ 先頭を無視：1行分
- ・ カラム名として解釈：2行分

設定後、画面の一番上の「名」の欄に名前を付けて【プロジェクト作成】をクリックする。複数シートも一気に読み込めるが、上記の設定では2つめ以降のシートの最初の行に、元の4行目5行目に入力されていたRequired(必須)などが記載されているので注意する。

#### 4.2 複数のカラムに分割

見出しを見ながら複数のデータを入れた列(カラム)を探して、以下の作業を繰り返す。

- ① 各見出しのプルダウンメニューから「カラムの編集」、「複数のカラムに分割…」をクリックする。
- ② 「カラムの分割をするには」で「文字で分割」を選択。
- ③ 区切り文字にパイプ (|) を入れて【OK】とする。

この作業により、各行に最大何個データが入力されているかを数えなくても、著者名やキーワードを複数のカラムに分割してくれる。それぞれの言語設定は、一括登録をした後に入力した方が自動で入力列を作成してくれるので手間が少ない。

#### 4.3 出力

画面右上の出力ボタンから、Excel形式でプロジェクトをエクスポートできる。最初に

付けたプロジェクト名がそのままファイル名となる。

エクスポートしたExcelファイルは、アイテムタイプが記載されていた1行目が削除され、2行目の内部キーと3行目のラベルがスペースで連結されて見出し行になっている。また、増設した列の見出しには「元の見出し(内部キーとラベル)」と1から始まる連番が付けられている。

### 5. Excelにて増設した見出しに連番をふる

元の一括登録用のメタデータに戻すためには、1行目を追加し、増設した列の見出しに0から始まる連番をふる必要がある。本学では以下の方法をとっているが、元のExcelファイルへ増えた列のみを挿入する方法もある。

- ① 先頭に3行追加し、1行目に元のファイルからアイテムタイプをコピーする。
- ② 見出しをすべてコピーする。
- ③ 別のシートにて右クリック、「形式を選択して貼り付け」を選び、行列を入れ替えて張り付ける。
- ④ 見出しを「内部キー」と「ラベル」と「連番」に3分割する。データの区切り位置機能を使う。区切り文字としてスペースを選び、表示形式をすべて文字列とする。
- ⑤ 連番が付けられた見出しに対して、連番マイナス1の数字を組み込む。  
例) D17セルに「=」キーワード["&B17-1&"].主題」と設定すると、B17セルに連番として1が入力されているので「キーワード[0].主題」が表示される。
- ⑥ オートフィル機能で0からの連番に振り直し、値のみを元の見出しに張り付ける。
- ⑦ 2列に分かれた見出しをコピーして、再び行列を入れ替えてデータが入っているシートの2行目3行目に張り付ける。項目のずれがないことを確認して、最初の見出しを削除する。
- ⑧ 必要なら増えた列の4行目5行目を設定

して、メタデータファイルを整える。

以上のように見出しを修正し、Excel形式のメタデータからタブ区切りのファイルを作成して一括登録をおこなう。

著者名やキーワードの言語は、登録したアイテムをエクスポートして作業する。一度登録をすると自動的に各項目の子項目の列が作成される。フィルターを使いながら言語を追記して、ふたたび一括登録をおこなう。

### 6. おわりに

以上のように、区切り記号さえ設定しておけばOpenRefineは自動的に最大数の列を増やしてくれる。Excelの区切り位置機能よりも非常に簡単である。しかし、この機能のためだけにOpenRefineをインストールして使い方を覚えるのは面倒だと思われる。

他にもデメリットとして、セルを分割してから出力するExcelファイルに見出しの修正が必要なこと、インポートやエクスポートを繰り返す必要があること、日本語のマニュアルが少ないことなどがあげられる。

それでもOpenRefineには、今回紹介できなかったが豊富な機能がある。抄録のなかに無意味な改行があったとしても、容易に削除することができる。また名前や出版社名に表記の揺れがあった場合でも抽出して修正することができる。知っておくとメタデータを扱う際の強力な手助けとなるはずだ。

この特集を機にOpenRefineの活用例について色々と情報交換ができればと思う。

(かじ・さちこ／安田女子大学図書館)

kaji-s@yasuda-u.ac.jp

## データスキル勉強会へのお誘い

田辺 浩介

大学図書館研究会の学術基盤整備研究グループでは、「目録とメタデータ」と「電子リソース管理」を中心に学術基盤整備について広く研究しています。

- ・ 活動は基本的にオンライン
- ・ 連絡や情報交換はチャットツールのSlackを使用
- ・ Zoomによるオンライン読書会を月2回、日曜日の朝9:00～10:00に開催
- ・ グループへの参加は大学図書館研究会の事務局組織担当まで

(<https://www.daitoken.com/aboutus/contact.html>)

図書館システムや機関リポジトリにおいて、書誌データはもちろんのこと、研究データの扱いまで求められる現代の大学図書館において、情報技術の重要性が高まっていることは論を待たないと思われまますが、では図書館員はどのようにして、そこで求められる情報技術を学べばよいのでしょうか。

情報技術を学ぶための教材は、インターネット・書籍を問わず大量に流通しています(大学の機関リポジトリにも、プログラミングや統計学の科目の教材が掲載されていま

す)。しかし、これらの教材は当然ながら「図書館員の業務でどのように使えるのか」という視点では制作されていないため、図書館で働く私たちがこれらの教材を使って学習を始めようとしても、なにから始めればよいのか、どのように業務の役にたつのかのイメージがわからず、続けることが難しいのではないかと思います。

そこで学術基盤整備研究グループでは2022年より、図書館員向けの情報技術の教育活動を行う非営利団体“Library Carpentry”(<https://librarycarpentry.org>)の提供する教材を用いて、オンライン勉強会を行っています。Library Carpentryの教材は、コマンドラインツールの操作方法や正規表現といった基礎知識から、PythonやRなどのプログラミング言語、GitやOpenRefineなどのデータ操作のためのツールの使い方などを、書誌データのような図書館の業務でなじみのある題材を用いて解説しています。これまでに本勉強会では、シェルスクリプトやOpenRefineの教材を学んできました。2024年4月より、新たにプログラミング言語Pythonの教材を用いて、データの集計や可視化の勉強会を始める予定です。みなさまの参加をお待ちしています。

(たなべ・こうすけ/物質・材料研究機構)

大学図書館研究会  
学術基盤整備研究グループ

ホーム

- ▼ オンライン読書会
- ▼ 全国大会分科会
- ▼ その他の活動

## 大学図書館研究会

### 学術基盤整備研究グループ

本グループは、目録・電子リソースなど、図書館における学術基盤についての研究を行うことを目的として設立されています。

#### お知らせ

- ・ 「第54回全国大会第6分科会」のページを作成しました。(2023年12月22日)
- ・ 活動方針を更新しました。(2022年9月17日)
- ・ 「大学図書館協議研究会」から「大学図書館研究会」への名称変更に伴い、ホームページのタイトルを「大学図書館研究会 学術基盤整備研究グループ」に変更しました。(2021年3月1日)
- ・ ホームの活動方針を更新しました。(2021年1月24日)
- ・ 「活動紹介」のページに「第28回大隈研オープンカレッジ『オンラインミーティング開催講座』」を追加しました。(2020年4月6日)
- ・ 「オンライン読書会」のページを作成しました。(2020年3月14日)
- ・ 学術基盤整備研究グループのホームページを作成しました。(2019年10月29日)

---

大学図書館研究会関西3地域グループ合同例会  
「ネットで探せない『書誌の書誌』の  
書籍化の裏話」開催報告

---

山上 朋宏

---

2024年1月27日（土）、関西3地域グループ合同例会「ネットで探せない『書誌の書誌』の書籍化の裏話」が開催された。本合同例会は毎年京都・大阪・兵庫の3つの地域グループが合同で催している行事であり、今回は京都地域グループが企画・運営を担当し実施した。

本合同例会の講師は2023年6月に刊行された『探すツール：図書館、出版、メディア書誌の書誌』の著者である伊藤民雄氏（実践女子大学図書館）である。まず、本書につながる慶應義塾大学大学院での研究、『インターネットで文献探索』各年版（日本図書館協会）等の執筆、データベース『図書館情報学文献目録』（BIBLIS PLUS）の構築について説明があった。

本書は2020年1月に『図書館情報学文献目録』構築時の作業が契機になり、2021年の夏には Libraries Unlimited の『Reference Sources in the Humanities Series』のような書誌の出版刊行を意識したそうである。2021年11月には日本図書館協会へ出版を打診し、2022年2月に出版が了承された。

執筆において苦労した点について伊藤氏は書誌の現物確認と収集を挙げていた。日本の書誌については勤務先の実践女子大学と近隣の都立多摩図書館、日野市立図書館、府中市立図書館で現物確認することができた。一方、欧米の書誌の所蔵は自館では1割程度だった。国内の所蔵は筑波大学と京都大学に集中しており、それぞれ訪問して現物確認を行った。アジアやロシアの書誌については購入代行サービスで資料の収集を行った。代金を支払ったのに届かなかった事例もあり、大変苦

労されたようである。また、調査の過程で Internet Archive、Google Books、HathiTrust に7～8割程度、電子化されていることが判明し、10年前だと本書は執筆できなかったかもしれないとのことである。

続いて書誌項目と執筆について解説があった。項目については天野敬太郎「図書館学書誌の書誌」（『図書館界』20(6)1969）の10項目にほぼ合わせ①領域②言語③種別④期間⑤配列⑥点数⑦解題⑧索引⑨所蔵情報⑩短評を選定した。これらの項目は各書誌の序文、凡例、本文、後書きから抽出し、外国語書誌の場合は汎用翻訳支援ツールPCOTを利用してテキスト化と翻訳を行った。書誌によっては関係性や継続性（同一出版社による正統な続編や、出版社・編者は異なるが序文などに続編を明示するもの）にも注意を払っている。

今後の展開としては、日本の図書館情報学の主要目録の空白期間を埋めることを目指していくとのことであった。具体的には、金沢文圃閣との雑誌復刻プロジェクトの付録として戦前戦後学校・児童図書館関係文献目録を発行予定であること、『図書館情報学文献目録』の拡充を進めていくことが表明された。

「書誌の書誌」としての本書のユニークさ、有用性はもちろん大学職員として勤務しながら研究・調査しまとめ上げる執筆過程の苦労話や新しい発見は興味深く、参加者からのアンケート結果も好評であった。

（やまがみ・ともひろ／

奈良女子大学学術情報センター）

## 議事要録

### 2023/2024年度 第6回常任委員会

日時：2024年3月20日（水祝）10:00-11:50

場所：Zoom

出席者（敬称略）：

呑海、赤澤、上村、有馬、小山、和知（以上、  
常任委員）

### 2023/2024年度 第3回全国委員会

日時：2024年3月24日（日）13:30-16:00

場所：Zoom

出席者（敬称略）：

小林[北海道地域]、加藤[千葉地域]、下山[東京地域]、中川[東海地域]、山上[京都地域]、吉田[大阪地域]、徳田[兵庫地域]、諏訪[広島地域]、柿原[九州地域]、楫[学術基盤整備研究]（以上、グループ推薦全国委員）、呑海、赤澤、有馬、上村、北川、小山（以上、常任委員）、青山、磯本（以上、常任（特定）委員）

◆議事の詳細は以下からご覧ください。

<https://www.daitoken.com/committee/>

## 組織通信

1-3月分（2024/1/1-2024/3/31）（敬称略）

### ○入会

九州地域グループ

玉岡 兼治

北海道地域グループ

松村 尚彦

### ○退会（2023/2024年度末、2023/9/1以降）

東京地域グループ

清水 滋文

東海地域グループ

前田 勝典

京都地域グループ

上山 卓也

小林 倫道（2022/2023末除籍者）

無所属

杉山 潤子

伊藤 淳

田中 綾子

以上、現勢336名（2022/2023末除籍者除く）

## 大図研出版物アクセス方法変更のお知らせ

大学図書館研究会のデジタル化出版物ダウンロード情報は、会報『大学の図書館』2024年1月号でも告知しましたように、2023年1月に送付したおはがきによる接続情報による運用を継続しているところです。

この、はがきで情報をお送りする運用を、大学図書館研究会の新年度である2024年7月以降、dtkMLを通じてお知らせする方法に変更しますのでお知らせします。

なお、現在の「おはがきによる接続情報」は、大学図書館研究会の本年度末である2024年6月まで、その運用を継続します。

本件の詳細は、事務局から dtkML に2024年4月6日にお送りしました、『《お知らせ》デジタル化出版物ダウンロード情報の通知手段変更のご案内』という件名のメールをご覧ください。

なお、お手元におはがきがない場合は、事務局組織担当 [soshiki@daitoken.com](mailto:soshiki@daitoken.com) までご連絡をお願い申し上げます。折り返し、接続情報を記載したはがきを郵送します。

(大学図書館研究会事務局)

## 会員情報調査を実施します

2024年3月24日(日)に開催されました、2023/2024年度第3回全国委員会で、今年度中に会員情報調査(悉皆調査)を実施することが審議のうえ、了承されました。

前回の会員情報調査は2019年5月に行いました。前回の会員情報調査で、会員情報が更新され、事務局会務がしやすくなりました。今後も会務を安定して行うためには、定期的な実施が必須であり、頻度は5年が妥当であると判断いたしました。

手段については前回は踏襲し、調査票を郵送し、回答はウェブフォームをメインとする使用する予定です。

皆さまのご協力をよろしくお願い申し上げます。

(事務局組織担当)

大学の図書館 第43巻第4号 (No.605) 2024年4月25日 (毎月25日発行) ISSN: 0286-6854  
編集・発行: 大学図書館研究会 年間予約購読料: 送料共6,000円

□大学図書館研究会出版部 (出版物購入・問い合わせ窓口)

〒305-0033 茨城県つくば市東新井10-1-111 マザータンク気付

E-mail: shuppan@daitoken.com

<出版物購入代金等振込先> 三菱UFJ銀行 越谷駅前支店 普通口座: 1403054 大学図書館研究会出版部

□大学図書館研究会事務局

〒105-0013 東京都港区浜松町2-2-15 浜松町ダイヤビル2F

E-mail: dtk\_office@daitoken.com

<会費振込先> ゆうちょ銀行 振替口座: 00190-2-79769 大学図書館研究会

## 大学図書館研究会第55回全国大会のご案内

第55回全国大会が、以下の日程と内容で開催予定です。詳細については、今後、大会Webサイト等で随時ご案内していきます。

全国大会は大図研の中心となる活動です。ついては、会員各位のご参加はもとより、皆様の廻りで大図研の活動や大学図書館を巡る状況に関心をお持ちの方に、お知らせいただけると幸いです。

開催日程: 2024年9月21日 (土) ~ 9月22日 (日)

開催方法: オンライン形式

プログラム:

9月21日 (土)

- ・会員総会
- ・研究発表・事例報告
- ・課題別分科会

第1分科会: 資料保存 第2分科会: 学術情報基盤 第3分科会: キャリア形成

- ・交流会

9月22日 (日)

- ・課題別分科会

第4分科会: 大学図書館史 第5分科会: 出版・流通 第6分科会: 利用者支援

- ・シンポジウム

## 第55回全国大会実行委員会へのお誘い

全国大会に参加するだけでなく、実行委員として一緒に大会を作り上げませんか? 全国のメンバーと新しい経験をできる機会になりますので、会員各位の積極的なご参加をお待ちしています。実行委員の業務分担は、実行委員長からのご案内 (dtkML[dtk:04680 第55回全国大会実行委員会へのお誘い]: 2024年4月4日送信) をご覧ください。

ご不明な点や参加のご希望は、全国大会実行委員会 taikai@daitoken.com までご連絡ください。

(全国大会実行委員会)