

OATF 問題提起

機関リポジトリで収集している
コンテンツは
今のままでよいのか？
ーライセンス問題ー

新潟大学学術情報部長
JPCOAR運営委員
JAIRO Cloud共同運営本部長

高橋菜奈子

問題提起のきっかけ

- 研究・イノベーション学会第39回年次学術大会 企画セッションのディスカッションから
 - 発表資料：<https://www.jsrpim-daigakukeiei.jp/post/20241105>
 - **CCライセンスが**ついていないものは**オープンアクセスとは言えない**という点は**全員**の共通理解
 - オープンアクセスはオープンサイエンス、オープンイノベーションに資するものであるべきという指摘
 - 図書館が膨大な労力をかけてコンテンツを収集しているのに、**CCライセンス**のついていないコンテンツを集めていて、結局、他のステークホルダーから、**機関リポジトリはダメだ**と言われてしまう、**というのは望ましい未来ではない**という危機感
- 新潟大学のOA加速化事業の取り組みから
 - AIに読み込ませようとしたら、PDFからテキストの抽出できない課題
 - そもそも**AIの機械学習**のために読み込ませてよいのでしたっけ？著者からの許諾が必要かどうか、という疑問

→文化庁のAIと著作権のガイドラインあり。第30条第4項の権利制限。
<https://www.bunka.go.jp/seisaku/chosakuken/aiandcopyright.html>

JPCOARのこれまでの取り組み

- コンテンツ流通促進作業部会
 - 西岡千文先生の調査（2020年7月2日時点のIRDBのデータ）によれば、機関リポジトリではライセンスの記載が18.82%にとどまる
 - **西岡千文「機関リポジトリにおける識別子・ライセンスの付与状況」**
<https://jpcoar.repo.nii.ac.jp/records/680>
 - 調査報告書を受けて、識別子についての取り組みは実施したが、（高橋が主査の時期には）ライセンス問題には踏み込めなかった。
- 杉田委員長からの助言
 - オープンアクセスの程度として、**libre OA**（CC等の自由なライセンス下で自由に利活用できる）、**gratis OA**（読めるだけ）の概念があります。このことについて、私たちに何ができるか、頭の体操をしてみたい。
 - JPCOAR作業部会員のSlackに #libre-oaのチャンネルを設置して議論を開始。
 - 機関リポジトリのコンテンツに**CCライセンスをつけるにはどうしたらいいのか？**という呼びかけに作業部会員の皆さんが意見交換。

Slackでの話題から (グリーンOAのライセンス)

- 著作権譲渡契約に基づき、出版者（著作権者）が通常の出版をしている場合（通常のグリーンOA）は、**リポジトリからCCで公開できるのか？**
- むしろ、Elsevierの場合はCC-BY-NC-NDと決まっているのでは？
 - 論文の出版後（エンバーゴ終了後）：
 - 購読論文の著者最終原稿を機関リポジトリで公開することができます。
 - 著者最終原稿にはCC-BY-NC-NDの記載が必要です。
- 他にもある？
 - EmeraldこちらはCC BY-NC
 - IOPscienceもCC BY-NC-ND
- **出版社のポリシーをまとめたサイト**はあるか？
- Shelpa Romeoに項目あり。
 - 例 <https://v2.sherpa.ac.uk/id/publication/32133>
 - ちなみに、Shelpa Romeoは、JISCのOpen Policy Finder になる予定。
<https://openpolicyfinder.jisc.ac.uk/>
- 和文誌の場合、**新SCPJに項目**はあるか？
- **新SCPJ (β版)**に「ライセンス情報（出版社版）」「ライセンス情報（著者最終稿）」という項目あり。

Slackの話題から (紀要論文のライセンス)

- **紀要は**わりと簡単にCCとつけることができるのではないかと。全CCのジャーナルになってくださいって**編集委員会**にかけあえばよい。
- その割にすすんでいないのは何故？
 - ライセンスとしてのCCがよくわかっていない
 - わかっていても、現状は無料で投稿も閲覧もできるのでCCBYとするインセンティブがない
 - 著者との関係性の中でCCを全て一律に決めるのは難しい
 - オープンアクセスおよびリポジトリを敬遠する意識の教員の反発が怖い。
 - CCライセンスに一度設定すると撤回することができないという性質。アウトアウトでとりあえず進めてしまうという方法がとりにくい。
- **リポジトリ担当** ⇔ (**編集委員会** ⇔ **著者**) という関係の中での交渉・支援が必要。
 - 紀要のリポジトリ登録の合意の中にCCの考え方が入っていたのかを、再び、紀要編集者と話をする (ついでDOIもつけてもらえらるともっとよい。)
- 一つの雑誌の中でCCがいろいろなバリエーションになった時に、リポジトリ側でも個別なものを受けとる必要あり。

Slackの話題から (ライセンスはどこに記入するのか)

- JPCOARスキーマで**CCを入力するところ**が、メタデータのアイテム単位とファイルごととに分かれているのでは？
- 現在の**JAIRO CloudのUI**ではファイルごとに設定されるのでは？
 - JCではWEKO2時代から（JPCOARスキーマ誕生前から）ファイルごとにライセンスを付与する機能があり
 - ファイルごとのライセンスのほうがUI的に目立っている
- ファイルごとにつけられる方のライセンス情報は、CiNii Researchに受け渡されない。**アイテム毎につけられたほうの権利情報**が受け渡される
- どの単位にメタデータを作成するのか
 - JAIRO CloudのUIとJPCOARスキーマガイドラインの記述を合わせるのに加えて、ハーベストとの整合性にも気を使う必要あり。
 - JPCOARスキーマでは、**ファイルによってライセンスが異なるようなら、別々のアイテム（メタデータ）として登録する**という議論をしたはず。
 - 論文アイテムに根拠データを含めて登録する運用をしているが、研究データに権利設定がされていない場合が多いため、アイテム単位にはCCは登録していない。
 - 研究データの「メタデータ共通項目」では、権利情報はアイテム単位で設定する想定。

Slackの話題から (AI機械学習は可か)

- 著作権保護された論文をAIの学習データとして使用できるのか？
- 日本国の著作権法では、生成AIが登場するより前に定められた**第30条4の2**にてAI学習は権利制限に該当。他国の著作権法では日本のようなAIに対する例外規定がない国も多い
 - 第30条4の2があるからといって何でもできるわけではなく、文化庁でAIと著作権について**ガイドライン**あり。
<https://www.bunka.go.jp/seisaku/chosakuken/aiandcopyright.html>
 - 文化庁のガイドライン上では「学習」と「生成」を分け、生成内容に依拠性が認められる場合著作権侵害に該当するという判断
- 11/20付のスカラリーキッチン、ライセンスとAIによる学習について、ネガティブな観点からの記事
"Guest Post: Supply Chain of Writing Fools"
<https://scholarlykitchen.sspnet.org/2024/11/20/guest-post-supply-chain-of-writing-fools/>
 - Taylor&FrancisがMSに学術論文をデータとして供与した件でもAIと学術流通というのが大きな話題
 - 日本の著作権法だと「学習」を拒む権利というものの立ち位置が不透明
 - そもそも何で学術出版社に「出版権」ではなく「著作権」を譲渡しないといけないの？

Slackの話題から (その他のライセンス)

- 機械学習の文脈だとText and Data Mining (TDM) Licensingについても理解が必要。
- Essential Guide to Open Access ([Publishers' Licensing Services](#) Rights and Licensing Hub
 - <https://rightsandlicensing.co.uk/resources/essential-guide-to-open-access/>
 - 出版社のRights Managementのイニシアチブあり。
 - 図書館もライセンス問題を勉強して対抗できるようにしておきたい。
- 研究データと著作権、ライセンスの関係性についても知りたい。



ということで、
有識者をお招きして公開勉強会を
開催することになりました。