Unpaywall Data feedのご紹介

2022年6月21日

iGroup Japan/iJapan株式会社
笠間　和喜

目次

1. 会社案内
2. Unpwaywall Datafeedの背景情報
3. Unpaywall Data feedのご紹介

1.会社案内
2.iGroup Asia Pacific 会社概要

- 1983年のスタート
- 30程度のグループ企業
- アジアでの独立系の学術出版流通の大手
- Library Businessを中心に活動
- Scholarly Publishingビジネスとの強い関わり

Sparc Japanの関係で2006年にNIIで講演したこともあります。

Libraryの再定義：情報提供サービスだけでなく**プロフェッショナルスキル獲得**及び**スキル開発のためのコアセンター**

Postコロナの時代予測：AIの時代：
今後、あらゆる仕事がAIにとって変わられてしまう、自動車の自動運転技術然り、プロフェッショナルなスキルを獲得及び開発することで、お金を稼ぐことができる。

1.会社案内

1.iJapan株式会社

2012年登記、事業開始

主要事業

1. オンデマンドPrint事業
    1. ACS （ジャーナル)のPrint On Demand
    2. AGU（ジャーナル)のPrint On Demand
2. 研究成果関連サービス
    1. 論文画像チェックサービス(imaChek)
    2. 研究成果プロモーションサービス（Ies Research /Science POD）
    3. 研究評価ツール(Scholytics）
3. 電子コンテンツ及び関連サービス販売
    1. RDATool Kit（図書館カタログ規則）の販売
    2. RemoteXs及びMyLOFT等のリモートアクセスサービス
    3. Intelex Past Masterの販売
    4. その他の電子コンテンツ
4. システム開発事業
    1. NIMSとのTDM, Data Repository　（台湾チーム及びUKの会社と協業）
    2. NIIとのWEKO3のLibraryツール開発　（UKの会社と協業）
    3. 国内メディカル出版系システム開発（現在開発中）　（台湾のチームと協業）

iGroupとの提携関係
開発部隊の共有

1.会社案内
2.OurResearch社

OurResearch

NPO法人
設立目的：研究をよりオープンにするためのツールを作成！
基本的には、ファンドからの資金提供を受けて、プロジェクト単位で完結しております。現状Unsubのみユーザからのご要望を確認して、最低限の費用をいただいて開発を継続させていただいている状況

Impactstory、oaDOIを公開
2016年11月21日 | 北米・中南米
毎年10月に世界各国で開催されるオープンアクセスウィークにおいて、オープンソースの研究者支援ツールImpactstoryはoaDOIを公開した。論文に付与される"doi.org/"を"oadoi.org/"に置き換えると、オープンバージョンの論文にリダイレクトするもの。

このツールは以下のデータソースをあたって、オープンバージョンの論文を探し出す。
Directory of Open Access Journals（DOAJ）で、OAジャーナルとして収録されているか
CrossRefのライセンスメタデータフィールドで、オープンライセンスとなっているか
ImpactStoryのDOIプレフィックスリストで、プレプリントリポジトリがあるか
DataCiteで、オープンデータセットであるか
BASE OA検索エンジンで、グリーンOA版があるか
BASEでOAか否か決定できない場合はリポジトリに直接？？
ジャーナルの無料PDFリンク

元々 https://oadoi.org/としてDOIを入力してOAを探すサービス
PMC, BASE, DOAJ, レポジトリ及び学術出版社のレコード

Our projects

unsub
A data dashboard to help libraries cancel subscription journal Big Deals.

unpaywall
A free database of open access scholarly articles, with an API and browser extension.

GET THE RESEARCH
An academic search engine for people outside academia.

アカデミア外の人々のための学術検索エンジン

depsy
An analytic platform to show the full impact of research software.

研究ソフトのインパクトを提示するための分析プラットフォーム

Impactstory Profiles
A researcher impact profile that highlights the impact of Open Science activities.

OpenScience活動のインパクトにハイライトを当てた研究者のプロファイル

paperbuzz.org
An open source of altmetrics data, based on Crossref Event Data.

CrossrefのEvent Dataをベースにした Altmericsデータのオープンソース

CiteAs[1]
An easy way for researchers to correctly cite the research software they use.

http://ourresearch.org/

UnpaywallはOAを探すブラウザの 拡張機能のサービスで有名です。
このサービスを行うことで、利用者は、論文単位で、OAを取得するとともに、版元は、利用者の分析とともにOAのDBを取得しUPDATE！

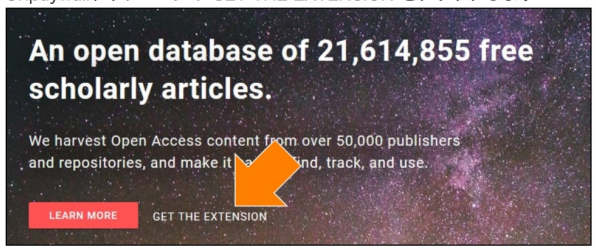
OurResearch社の実績：
Unpaywall-千葉大学附属図書館での案内

Unpaywall

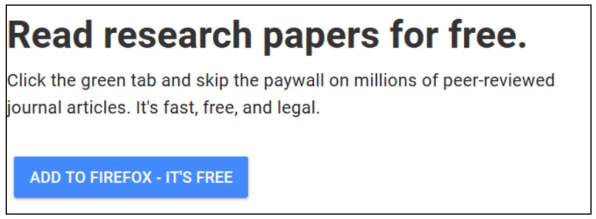
対応しているWebブラウザ：Chrome、Firefox

【利用手順】※ここではFirefoxを例に説明します。

1. Unpaywall [外部リンク]にアクセスします
2. Unpaywallトップページの"GET THE EXTENSION"をクリックします



3. "ADD TO FIREFOX"ボタンをクリックします



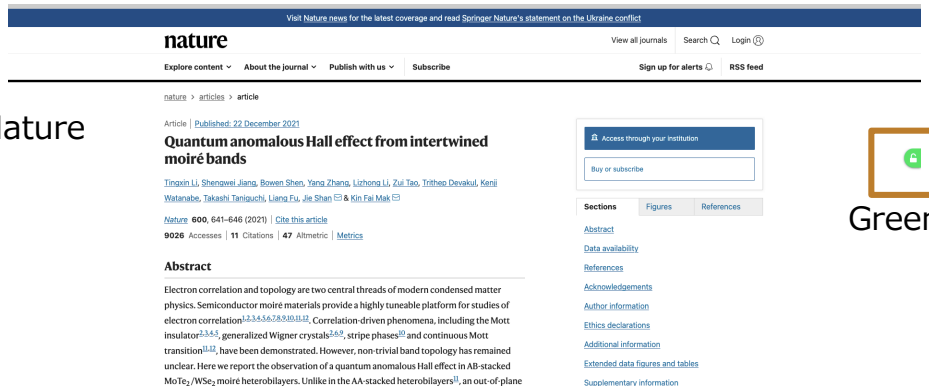
4. "Add to Firefox"ボタンをクリックすると使用できるようになります

参考URL: https://www.ll.chiba-u.jp/openAccess.html

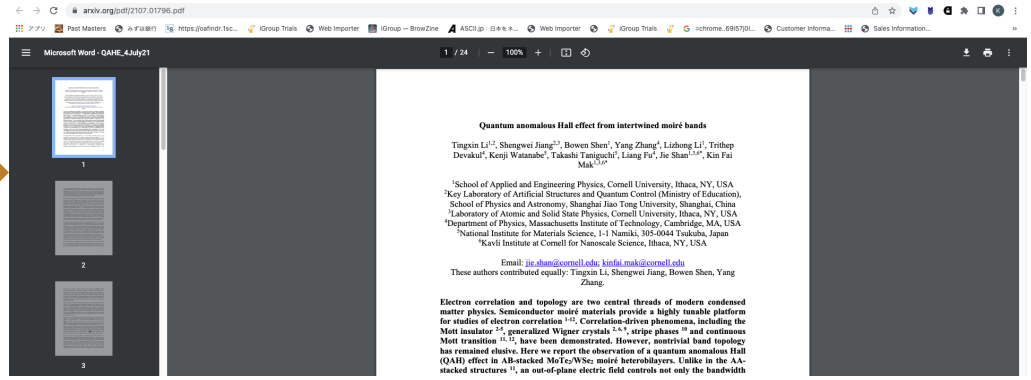
OurResearch社の実績

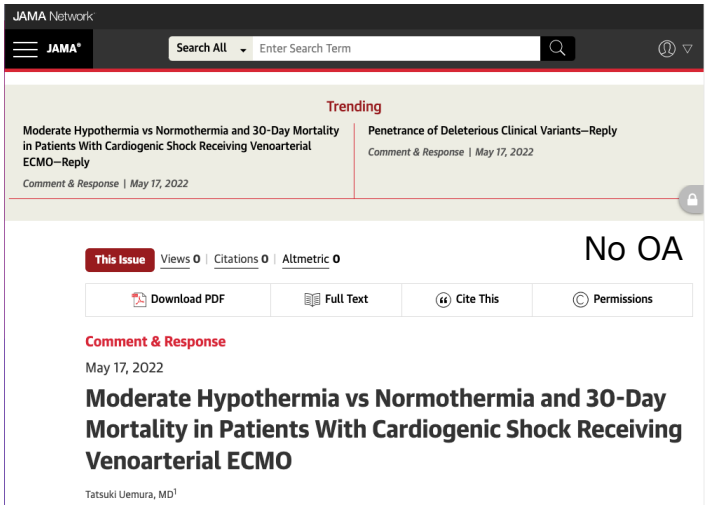
unpaywallってなに？

Nature

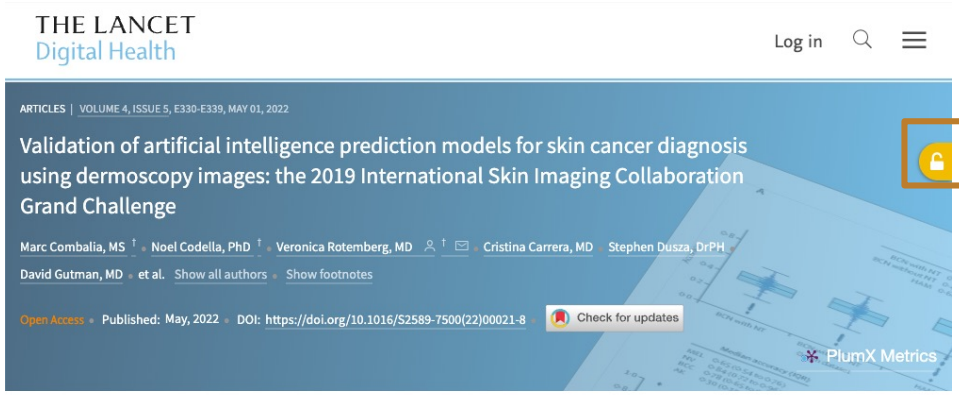


Green OA

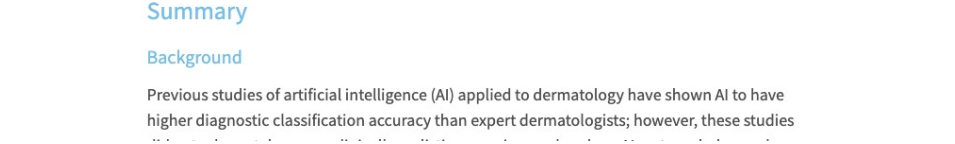




No OA



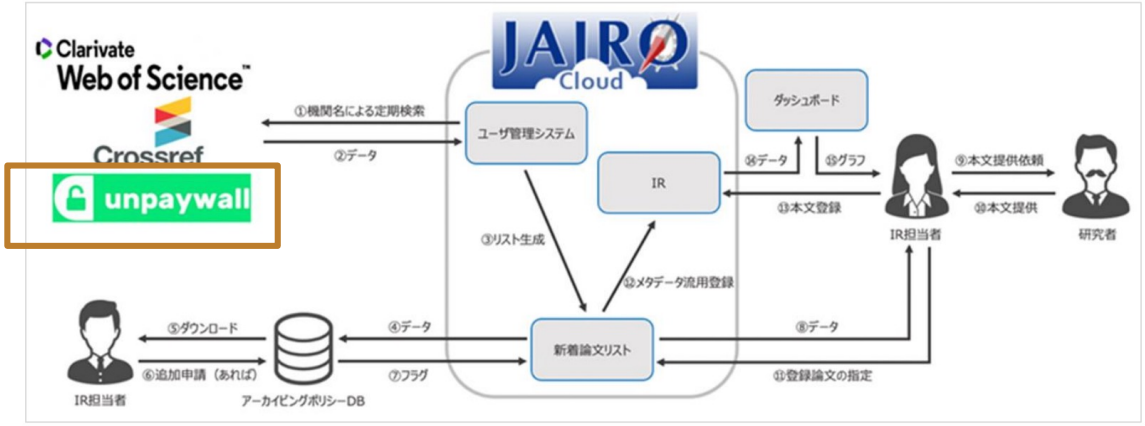
Gold OA



OurResearch社の実績：
Unpaywall-Jairo cloudとの連携

NII RCOS
Research Center for Open Science and Data Platform
National Institute of Informatics

グリーンOA支援システム



- 図書館員による作業を軽減化するためにシステムの開発と普及（NII＋JPCOAR）
- 実証実験
  - 第1段階
    2020年9月～2020年11月
    参加機関：1（筑波大学）
  - 第2段階
    2021年1月～2021年3月
    参加機関：4（大阪大学、神戸大学、九州大学、一橋大学）
  - 第3段階
    2021年11月～12月（予定）
    参加機関：5（横浜国立大学、新潟大学、信州大学、福井大学、熊本大学 ）

図書館総合展
「機関リポジトリからみた管理・検索基盤」
2021年11月9日
国立情報学研究所
オープンサイエンス基盤研究センター
尾城 孝一

https://www.nii.ac.jp/event/upload/libfair2021_forum1_1.pdf

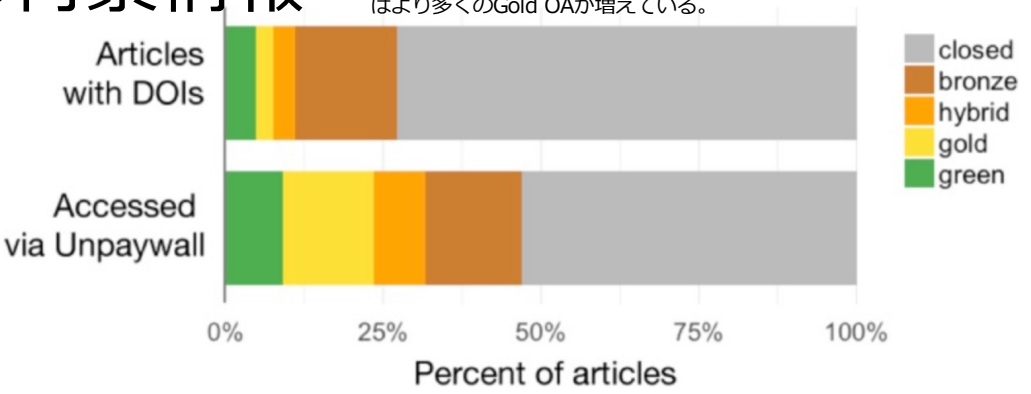
NII様のJAIRO CloudもUnpaywall及び
OA発展にご協力いただいている状況かと思います。

2. Unpwaywall Datafeedの背景情報

1.OpenAccess論文の台頭

1.OpenAccess化率

なぜGold OAが増えたか？
上の図の通り、GoldOAの定義としてDOAJ収録誌を対象にしていたが、DOAJ以外にもCreative Commons等で確認されたGoldOA誌がたくさんあったため、下図ではより多くのGold OAが増えている。

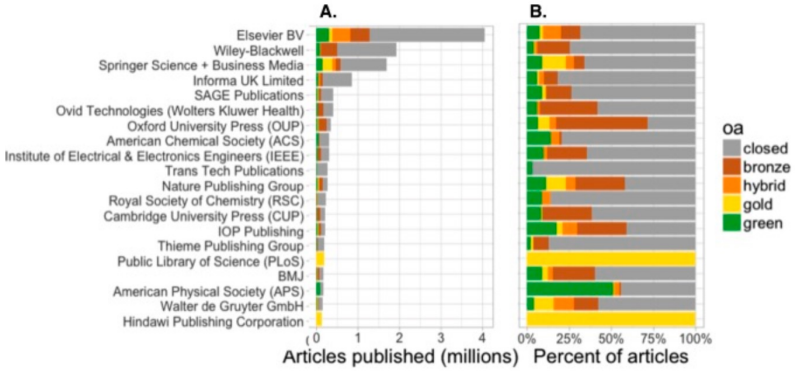


PMC full text: PeerJ. 2018; 6: e4375.
Published online 2018 Feb 13. doi: 10.7717/peerj.4375

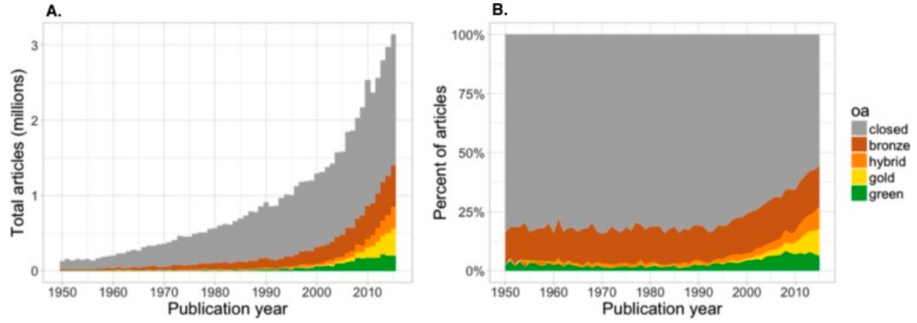
Percent of articles by OA status, Crossref-DOIs sample vs Unpaywall-DOIs sample.

Figure 2

Figure 3



Number (A) and proportion (B) of articles with OA copies, by publisher, for the 20 most prolific publishers. Based on sample of 27,894 Crossref DOI-assigned articles published between 2009–2015.

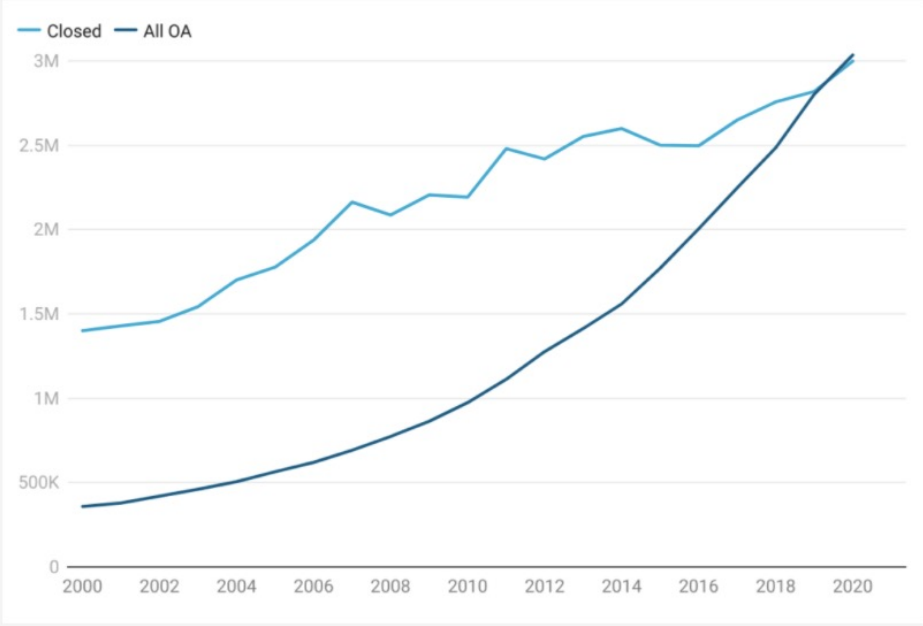


Number of articles (A) and proportion of articles (B) with OA copies, estimated based on a random sample of 100,000 articles with Crossref DOIs.

増えるOA論文
隠れOA論文（USの補助金関連の論文
1年間のエンバーゴ（出荷停止期間））

2. Unpwaywall Datafeedの背景情報

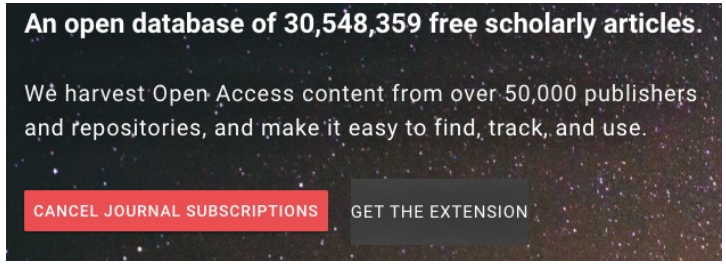
1.OpenAccess論文の台頭



2020年ついに、ClosedよりOAの方が
論文数が多くなった？

Open Access surpasses subscription publication globally for the first time

https://www.dimensions.ai/blog/open-access-surpasses-subscription-publication-globally-for-the-first-time/



↑
2021年11月17日時点の論文数
https://unpaywall.org/

2. Unpwaywall Datafeedの背景情報

1.OpenAccess論文の台頭

2.OA論文の将来

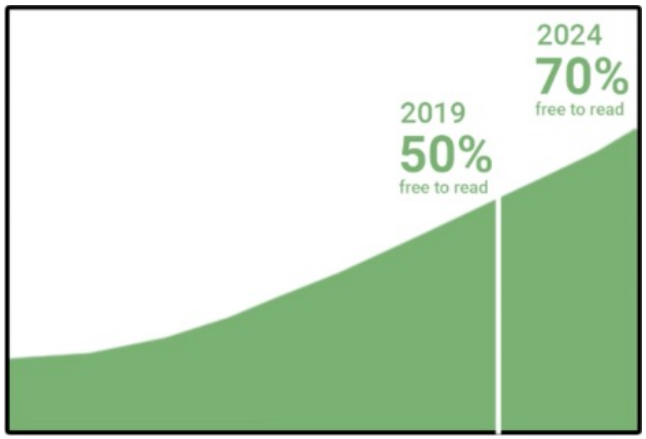
OA and Subscription Journalsのトレンド-予測 1

50%

of article views could be fulfilled by OA right now..

70%

will be in five years.

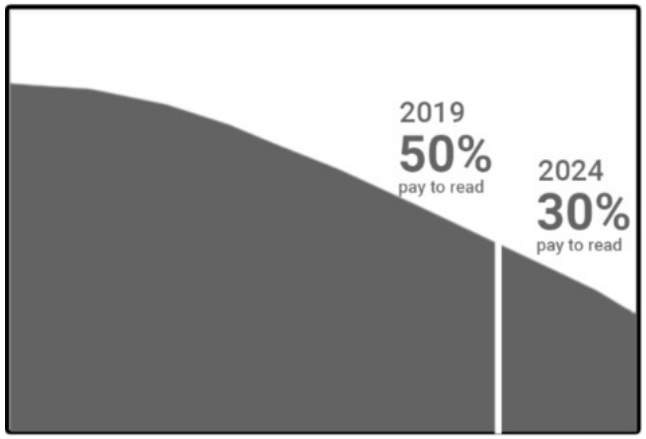


50%

of article views must be purchased now.

30%

must be purchased in five years.



https://vimeo.com/427553921

Percent of views, by OA type:

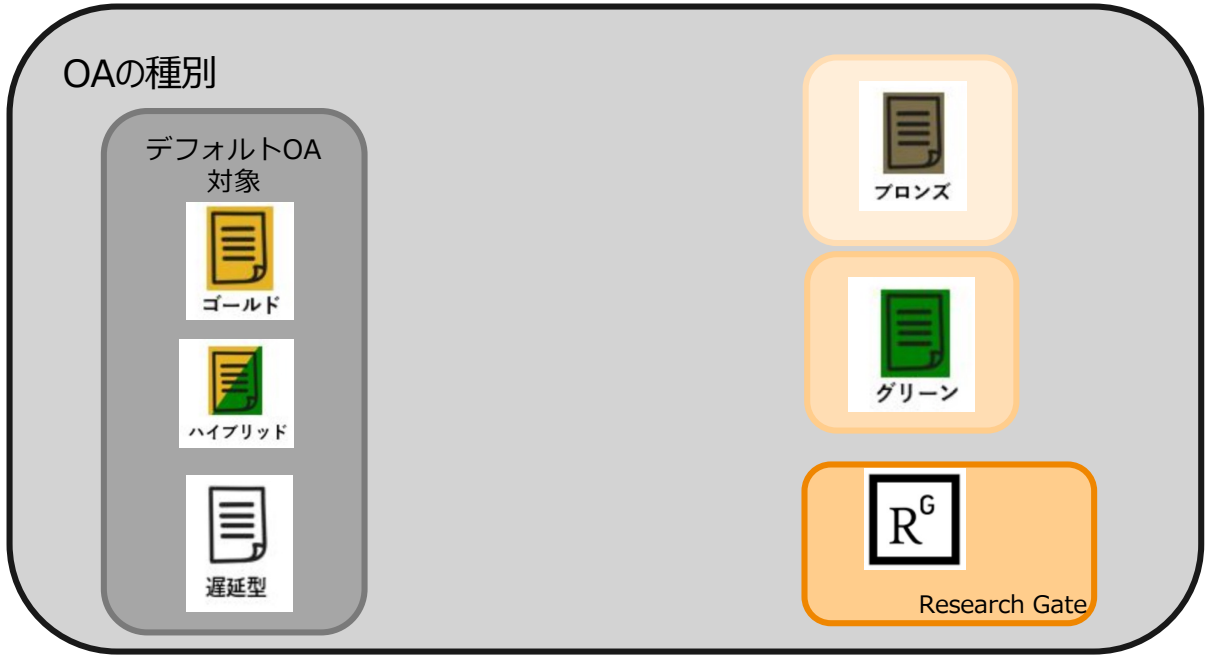
year	green	gold	hybrid	immediate_bronze	delayed_bronze	closed	all OA
2010	6	6	3	6	8	71	29
2019	15	20	7	5	6	48	52
2025	19	33	10	4	4	30	70

doi: https://doi.org/10.1101/795310

2. Unpwaywall Datafeedの背景情報

1.OpenAccess論文の台頭

3.OA論文の種別



慶応義塾大学
倉田先生
https://www.mext.go.jp/kaigisiryo/content/000030344.pdf

OAの対象

Gold OA
＋
HyBride OA
＋
遅延OA
＋
Bronze OA
＋
Green OA
＋
Research Gate OA

GoldOA等　画像利用：
https://libguides.lib.keio.ac.jp/oa/what
Wikipedia

3. Unpaywall Data feedのご紹介

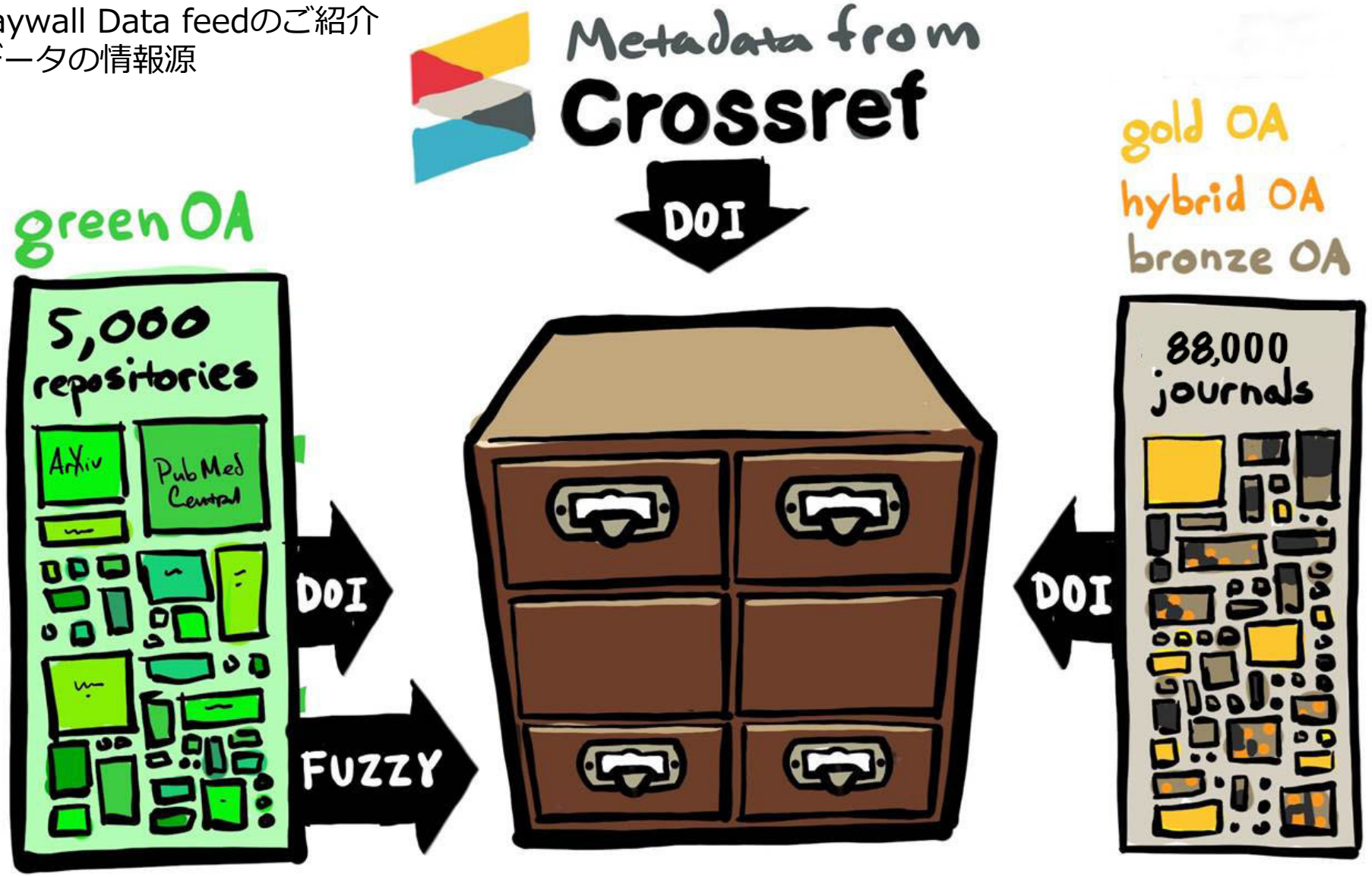
unpaywall

The Unpaywall Data Feed

（unpaywallのサービスを構築するにあたり、作成されたOAデータ及びメタデータ）

A complete, open, and legal index of open access.

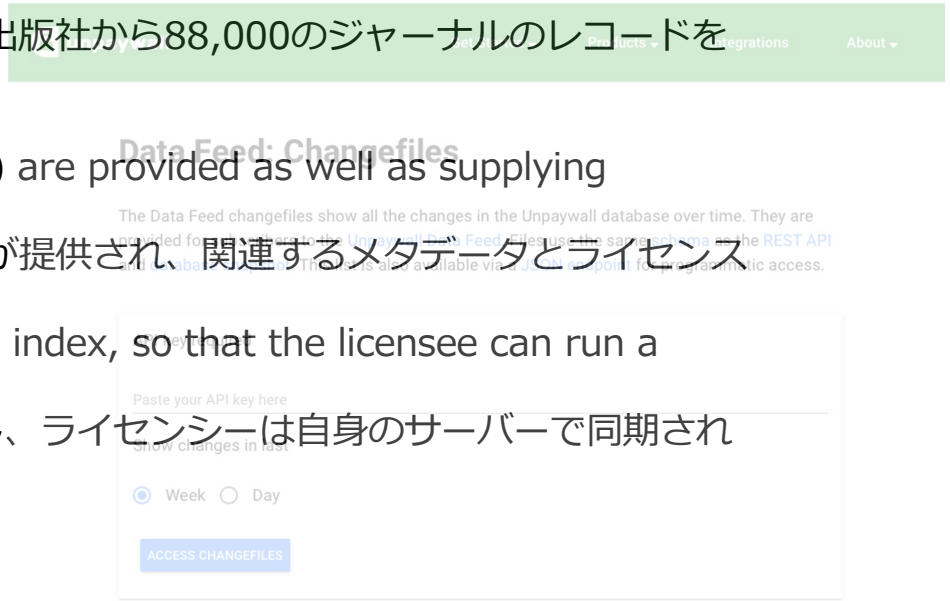
3. Unpaywall Data feedのご紹介
1. OAデータの情報源



3. Unpaywall Data feedのご紹介
2. OAデータの諸元

The gold standard for discovering Open Access academic research.
オープンアクセスの学術研究を発見するためのゴールドスタンダードです。
Tracks the status of over 135 million articles, about 35 million of which are Open Access – growing daily
1億3500万件以上の論文の状況を追跡し、そのうち約3500万件はオープンアクセスで、日々増加しています。
The Data Feed draws from over 5,000 repositories, and has records for 88,000 journals from 14,000 publishers.
データフィードは、5,000以上のリポジトリから取得し、14,000の出版社から88,000のジャーナルのレコードを有しています。

For any given DOI, all the available OA copies (if there are any) are provided as well as supplying relevant metadata and license information.
任意のDOIについて、利用可能なすべてのOAコピー（もしあれば）が提供され、関連するメタデータとライセンス情報も提供されます。
The Data Feed product provides daily changefile updates of the index, so that the licensee can run a synced copy on their own servers.
Data Feed製品は、インデックスの毎日の変更ファイル更新を提供し、ライセンシーは自身のサーバーで同期されたコピーを実行することができます。



3. Unpaywall Data feedのご紹介
2. OAデータの諸元

The Unpaywall Data Feed is available in 3 main ways – two are free. They include:
Unpaywallのデータフィードは、主に3つの方法で利用できます。 そのうちの2つは無料です。

1. **Data Feed Snapshot** - comes out twice per year – available on the website
Data Feed Snapshot - 年に2回発行され、ウェブサイトから入手できます。
2. **REST API** - provides up to 100,000 API calls per day – find instructions on the website
REST API - 1日あたり最大100,000のAPIコールを提供します - 詳細はウェブサイトをご覧ください。
3. **Paid Subscription** to the Data Feed
- Provides real-time updates with daily changefiles available
- Pricing is based on your intended use, your size
- We can provide special pricing as well if you wish to use the REST API but go over the free threshold of 100,000 API calls per day

データフィードの有料購読
・リアルタイムの更新を提供し、毎日のチェンジファイルが利用可能です。
・価格は、お客様の使用目的、規模に応じて決定されます。
・REST APIを使用したいが、1日あたり10万回のAPIコールが必要な場合は、特別価格で提供することが可能です。

3. Unpaywall Data feedのご紹介
2. OAデータの諸元

DOIがある論文を対象
135,000,000+ meta data
+articles status check
35,000,000+ OA articles.
14,000 Publishers
88,000ジャーナル対象
5,000以上のレポジトリからの情報

OAデータ種別内訳		
種別	論文数	%
Medicine(医学)	6,300,000	18%
Chrmistry(化学)	3,850,000	11%
Computer science(コンピュータ科学)	3,850,000	11%
Materials science(材料科学)	2,800,000	8%
biology(生物学)	2,450,000	7%
Psychology(心理学)	2,100,000	6%
Political Science(政治学)	2,100,000	6%
Physics(物理学)	1,750,000	5%
Art(芸術)	1,400,000	4%
Business(ビジネス)	1,400,000	4%
Sociology(社会学)	1,400,000	4%
Mathematics(数学)	1,050,000	3%
Environmental Science(環境科学)	1,050,000	3%
Philosophy(哲学)	1,050,000	3%
Geography(地理学)	700,000	2%
History(歴史)	700,000	2%
Geology(地質学)	350,000	1%
Economics(経済学)	350,000	1%
Engineering(工学)	350,000	1%
total	35,000,000	100%

3. Unpaywall Data feedのご紹介
3. 事例-1 https://scite.ai/

sciteは、これまでに発表されたすべての研究論文の情報を整理することを使命としています。私たちは 研究論文の全文から、他の論文への言及がある文章を抽出することによって を抽出します。これらの文章を分離し、どの論文を引用しているかを特定することで、あらゆる研究論文とその論文がどのように扱われているかを知ることができます。どのような研究論文とその主張が、どのような文献に引用されているのかを知ることができます。ある論文が
ある論文が100回引用されたことを知る代わりに、その100の論文のそれぞれでその論文について何が語られたかを一度に知ることができます。を一度に見ることができます。

Unpaywallの支援方法
Unpaywallは、どの論文がオープンアクセスで、どのようにダウンロードできるかという情報を含むデータベースを維持します。他のメタデータと一緒に）ダウンロードすることができます。データベースのスナップショットに加え、次のような機能を提供しています。データフィードを提供し、指定されたウィンドウ内でデータベースへのあらゆる変更を伝える変更ファイルに、プログラムから頻繁にアクセスできるようにしています。データフィードを提供し、指定された時間内にデータベースが変更された場合、変更ファイルにプログラムから頻繁にアクセスできるようにしています。
そのため、新しい論文が出版されたり、既存の論文の禁書が解除されてオープンアクセスになったりすると、Unpaywallのチェンジファイルが更新されます。Unpaywallの変更ファイルにより、オープンアクセス・カタログ全体の最新情報を入手することができます。

データフィードが特に優れているのは、変更ファイルにアクセスする方法に柔軟性を持たせている点です。毎日、毎週、変更ファイルを消費するだけでなく、次のようなタイミングで変更ファイルを検索することもできます。また、（取り込みシステムの技術的な問題で）何か見逃したものがあれば、以前の変更ファイルに戻ることもできます。を取得することができます。）
オープンアクセスのコンテンツで特に問題となるのは、URLの破損です。ある特定の論文が 例えば、SSLの問題やホストの気まぐれによって、あるURLで特定の論文にアクセスできなくなることがあります。リンクが機能しないケースはありますが リンクが機能しない場合もありますが、そのような場合、より新しい、機能するURLに更新され、その後の変更ファイルを通じて提供されることがよくあります。これは、私たちの安心感を大きく向上させるものです。

3. Unpaywall Data feedのご紹介
3. 事例-2 https://www.scholarcy.com/
効率よく論文を読み込むためのサマリー作成サービス

状況:
研究発表のメディアとしてのPDFは、オフラインで読め、注釈を付けられ、どのデバイスでも読めるという多くの強みを備えています。共有でき、あらゆるデバイスで読むことができる。しかし、PDFには多くの問題があります。また、含まれる情報は、専門家でない人を排除するような背景知識を前提としていることが多いため、様々な視覚的・認知的ニーズがあります。を想定していることが多いこと、また、インタラクティブでないことが多いため、簡単に文献を追ったり、著者の主張を検証したりできないことです。そのため、参考文献を調べたり、著者の主張を検証したりすることができません。

課題:
私たちは、プレプリントを含む研究資料を、専門家でなくてもアクセスできるインタラクティブな体験にまとめたいと考えました。また、すべての引用をクリックできるようにすることで、出典や主張をより簡単に検証できるようにしました。また、すべての引用文献をクリックできるようにして、出典や主張をより簡単に確認できるようにしたいと考えていました。しかし、現在でもほとんどの著者は、参考文献リストにDOIを記載していません。しかし、現在でもほとんどの著者は、参考文献リストにDOIを提供しておらず、それは記録されたバージョンへのリンクを容易にするものです。たとえDOIを提供していたとしても、理想を言えば、私たちは
自由に利用できるバージョンの参考文献に直接リンクできることが理想です。

解決策
研究用PDFをインタラクティブなHTMLに変換するために、カスタム機械学習モデルとルールエンジンを構築しました。Wikipediaへのリンク、重要な手法や結果の自動ハイライトなど、専門家でない方にも役立つ機能を備えています。Wikipediaへのリンク、主要な手法や研究結果の自動ハイライト、さまざまなタイプの要約（上位5つのハイライトと、より長い叙述的な要約）など、専門家でない人にも役立つ機能を備えています。
また、文献を抽出する解析エンジンを構築し、CrossRef APIを使用してそのDOIを取得できるようにしました。しかし、ケーキの上のアイシングはUnpaywall APIで、各参考文献について、公開されているかどうかをチェックすることができました。各文献について、オープンアクセス版があるかどうかを確認し、引用された各文献のフルテキストに直接リンクできるようにしたのです。引用されたソースのフルテキストに直接リンクすることができました。

結果
添付のビデオはその結果を示しています：私たちのChrome Extensionは、ブラウザで現在のPDFを処理することができ 引用をクリックできるインタラクティブなフラッシュカードに変えることができます。これと、私たちの有料会員制製品。Scholarcy Libraryは、現在、世界中に何千人ものユーザーを持ち、多くのお客様に満足していただいています。ALPSP 2020アワードと2020 Vesalius Innovationアワードの最終選考に残りました。

3. Unpaywall Data feedのご紹介

3. 事例-3 https://www.tripdatabase.com/

Trip medical database 臨床研究エビデンスを見つけるためのツール

課題

顧客を維持し、新規顧客を獲得することは、最良の状態でも困難です。Jonは ジョンさんは、顧客の価値を高め、製品を差別化する方法を探していました。さらに、元研究員として を見つけること、そしてそれを購入することは、組織にとって費用と時間のかかることであることを知っていました。この問題を解決し、同時にTrip Databaseを強化する方法はないだろうか？

解決策

Trip Databaseは、Unpaywall Data Feedを統合し、製品の機能性を向上させることで、次の2点を解決しました。製品に付加価値を与え、研究者にオープンアクセス研究へのアクセスを提供する必要があります。へのアクセスを提供することができます。Tripは、フルテキストへのリンクに加え、テキストマイニングを使用して、システマティックレビューという追加コンテンツを特定しました。システマティックレビューを追加しました。

結果

お客様からのフィードバックは、非常にポジティブなものでした。この機能は、時間とリソースを節約できる優れた付加価値であると 時間やリソースを節約することができます。この機能により、Trip Database の製品に新たなレイヤーが生まれました。製品に付加価値を与えることができました。

3. Unpaywall Data feedのご紹介
3. 事例-4 Swiss National Science Foundation (SNSF)
http://www.snsf-oa-check.ch/

状況
オープンアクセス（OA）研究は、政府がプランSのような動きでそれを義務化し、科学や学術界の多くがそれを推進しているため、日々ますます支持を集めています。この動きは、論文の閲覧が有料であった過去からの大きな変化であり、研究者や著者は、それが自分たちや自分たちのいる分野にどのような影響を与えるのかを理解することが重要です。スイス国立科学財団（SNSF）は、スイスにおけるオープンアクセスの推進役であり、この目的のために、学術研究者や著者は、OA化された学術論文の割合やそれらの論文の状態（ゴールド、グリーン、ハイブリッド、ブロンズ）を分析できるツールを設計しました。
1952年に設立され、スイスのベルンに本部を置くスイス国立科学財団は、科学研究支援組織として設立された。SNSFは、スイス連邦から研究資金の援助とスイスに住む若手科学者の育成を委任されている。SNSFの主な目標は、質の高い研究および研究者の育成を支援することです。
知識の普及を支援し、研究の価値を示すことです。
課題
増え続けるオープンアクセス論文を把握することは、地域レベルでも世界レベルでも非常に困難です。を把握することは、個々の研究者や著者はもちろんのこと、政府機関にとっても、地域や世界レベルで非常に困難なこと これは 困難な課題ですが、解決策がないわけではありません。これまで、このような個人を支援するツールはありませんでした。どのような著作物がOA化されているか、そのOA化の状況はどうか、どのようなリポジトリにあるか、どのような論文がOA化されうるか、などを常に把握できるツールはこれまでありませんでした。どのような論文をOA化することができるのか。
解決策
スイスを拠点とする研究者、著者、学術関係者を支援するために、SNSFはプロトタイプのツールを構築することを決定しました。ポータルを構築することにしました。このポータルには、オープンアクセス論文を追跡するメタデータが大量に含まれています。データを取り込む 5,000以上のデータベースで1億2600万以上の学術論文を追跡するOur ResearchのUnpaywall Data FeedからのデータとDimensionsからのデータを組み込んでいます。このツールは、5,000以上のデータベースで1億2,600万件以上の学術論文を追跡するOur ResearchのUnpaywall Data FeedとDimensionsのデータを組み込んでおり、スイス在住の研究者や著者が このツールは、スイスを拠点とする研究者や著者が自分の作品のOA状況を理解し、追跡できるようにするものです。
最終的な目標は、Dimensionsとともにこのツールの開発を継続し、Dimensionsが提供するサービスの一部となることを目指すことです。最終的な目標は、Dimensionsが提供するサービスの一部として、このツールの開発を継続することです。
結果
このツールは現在稼働中で、現在スイス在住の著者と研究者のみに、著作物の管理と追跡を行う新しい方法を提供しています。を管理・追跡する新しい方法を提供しています。このツールは、著者や研究者に新しい をリアルタイムで管理する新しい方法を提供します。

3. Unpaywall Data feedのご紹介
4. 価格情報

スタートアップ会社　$1500-$3000: 1$135円計算だと、20万強-40万強
出版社　＄100000　同じく、135万円

Unpaywall datafeedについての技術情報は以下のサイトにありますのでご参考にしてください。

https://unpaywall.org/products/data-feed

終わりに：

基本的には、どのようにOA論文及びOA論文のビッグデータを御社のサービス取り込むかにより、御社のシステムの付加価値が変わってくるかと思います。
ご存じの通り、引用や利用に関するデータは日々更新しており、業績管理についても提出して終わりという訳ではなく、日々自動的に更新されてゆくかと思います。
この機会に、OA及びCloseのビッグデータを如何に御社の業績管理システムと連結して、サービス提供してゆくかということについてご検討ください。

また、このunpaywallのDatafeedを作成している、OurResearch社では、OAとClosedのデータを統合して、OpenAlexというデータベース(https://openalex.org/)を構築しております。これはMicrosoft Academic Graphというサービスの後継にあたるもので、今後Unpaywall Datafeedと同様に、API Keyを販売してゆくと聞いております。

You're already using Unpaywall if you use:

- Web of Science
- Scopus
- ProQuest Summon
- OCLC WorldCat
- EBSCO EDS
- DS Dimensions
- Lens
- etc..

Thank You &
Let us know if you would like trial access