

SPARC Japan ニュースレターでは、各回セミナーの報告に講演や全体議論を書き起こしたドキュメントを加え、さらにそのほかの SPARC Japan の活動をご紹介します。

## CONTENTS

### ■ SPARC Japan 活動報告

SPARC Japan 運営委員会  
SCOAP<sup>3</sup> 支援

### ■ SPARC Japan セミナー報告

概要  
参加者から  
企画後記  
ドキュメント  
(講演・全体議論)

### ■ SPARC Japan 活動報告



#### SPARC Japan 運営委員会

SPARC Japan 運営委員会の会議資料をウェブサイトで公開しています。

<http://www.nii.ac.jp/sparc/about/committee/>

#### SCOAP<sup>3</sup> 支援

第5期 SPARC 事業の達成目標の一つとして「国際的な OA イニシアティブとの協調」および「学術情報流通にかかわるアドボカシー活動」を掲げ、その具体的活動として SCOAP<sup>3</sup> (スコープ・スリー) を支援しており、日本参加機関の取りまとめと広報活動を行っています。

SCOAP<sup>3</sup> (Sponsoring Consortium for Open Access Publishing in Particle Physics) は、スイスの CERN (欧州原子核研究機構) が中心となり高エネルギー物理学分野の査読付きジャーナル論文のオープンアクセス化を実現することを目的とした国際連携プロジェクトで、日本では高エネルギー加速器研究機構、国公立大学図書館協力委員会、および国立情報学研究所 (NII) が 2011 年に共同で関心表明 EoI に署名し、NII が National Contact Point となりました。

フェーズ 1 (2014 年～2016 年) における日本の参加機関は 34 機関で、フェーズ 2 (2017 年～2019 年) は 2017 年に 40 機関、さらに 2018 年からアメリカ物理学会が対象誌として加わったことをうけ、67 機関が参加を表明しています。「大学図書館と国立情報学研究所との連携・協力推進会議」のもとに設置されるタスクフォースでは、拠出金額の算定方法を定めるほか、参加意向と拠出金額の確認などを行うとともに、物理学研究者向けにも広報活動を行っています。また日本参加機関の拠出金は毎年 NII が取りまとめて CERN に支払い、各参加機関は NII に支払っています。

詳しい内容は Web サイトをご覧ください。(<http://www.nii.ac.jp/sparc/scoap3/>)

## ■ SPARC Japan セミナー報告



### 第2回 SPARC Japan セミナー2017 (オープンアクセス・サミット2017) 「プレプリントとオープンアクセス」

2017年10月30日(月) 国立情報学研究所 12F 会議室 参加者: 63名

本セミナーでは、オープンアクセスの展開において不可欠な要素であるプレプリントサーバについて取り上げました。プレプリントサーバの機能と変遷を確認しつつ、研究者と出版者からの発表を交えて、研究推進への寄与、持続性のあるビジネスモデルの構築、質の確保などの将来の課題を含めて議論しました。

次ページ以降に、当日参加者のコメント(抜粋)、企画後記およびドキュメント全文(再掲)を掲載しています。その他の情報は SPARC Japan の Web サイトをご覧ください。(<http://www.nii.ac.jp/sparc/event/2017/20171030.html>)

### 概要



オープンアクセスは、主にリポジトリを活用するグリーン OA と、ジャーナルを無料化するゴールド OA と呼ばれる方法のいずれかによって実現されると考えられている。arXiv を嚆矢とする分野別プレプリントサーバはこれまでリポジトリの一種別として位置付けられ、かつ、査読の前の草稿であることが原則であるため、研究成果の商業的流通への影響が少ない存在として理解されてきた。しかし、なかなかプレプリント流通が普及しなかった化学、生命科学の分野でも近年急速にプレプリント掲載数が増加する傾向が顕著となり、有力な学会誌をもつ国際学会(ChemRxiv の ACS 他)や商業出版者(BioRx の Elsevier)がプレプリントサービスの提供に参入してきている。

さらに、これらのプラットフォームはデータ公開を含めて設計されており、オープンアクセスの展開におけるプレプリントサーバの位置づけは大きく変化しつつあるといつてよい。

今年度の第2回目の SPARC Japan セミナーでは、プレプリントサーバの機能、運営の約30年の変遷を振り返りつつ、この変化の渦中にある研究者からの提言、さらに研究者ネットワークの商業出版者にとっての意義を論じ、研究推進への寄与、持続性のあるビジネスモデルの構築、質の確保などの将来の課題を含めてこれからのオープンアクセスを展望する。

第2回 SPARC JAPAN セミナー2017  
プレプリントとオープンアクセス

OPEN International ACCESS WEEK  
"Open in Order to ..."

日 時: 平成29年10月30日(月) 11:00-16:40  
場 所: 国立情報学研究所 12階 1208,1210会議室

プログラム

- arXiv.org の次世代システムの公開と戦略  
引原 隆士 (京都大学図書館機構長/arXiv.org 会費コンソーシアム代表)
- 学術情報共有とオープンアクセスの未来  
Gregg Gordon (Managing Director, SSRN) ※遠隔通訳付き
- 化学分野におけるプレプリントの位置付け・課題等について  
生長 幸之助 (東京大学大学院薬学系研究科/化学ポータルサイトChem-Station副代表)
- 生命科学分野におけるプレプリントの位置付けや経験について、統合TVについて  
小野 浩雅 (情報・システム研究機構 ライフサイエンス統合データベースセンター)
- 全体議論  
(モデレーター)  
坊慶 秀雅 (情報・システム研究機構 ライフサイエンス統合データベースセンター)

動画中継予定 URL: <https://www.nii.ac.jp/sparc/event/>  
お問い合わせ先: 国立情報学研究所 SPARC担当 E-mail: [co\\_sparc\\_mail@nii.ac.jp](mailto:co_sparc_mail@nii.ac.jp)

主催: SPARC Japan (国立情報学研究所)

NII 国立情報学研究所 SPARC\*Japan OPEN International ACCESS WEEK

## 参加者から

(大学/図書館関係)

・プレプリントの価値が認められている分野がある、そしてこれから最も存在価値が上がりそうということをリアルに感じられました。

・arXiv と SSRN の話の対比は非常に興味深い。持続可能性についてさらに議論を聞きたい。

(企業/図書館関係)

・図書館側からすると、プレプリントサーバの存在は、まだ注目できていない気がします。今後も、アンテナを張っていかないと、研究者の方からのレファレンスに答えられなさそうです。

(その他/図書館関係)

・最近の SPARC セミナーは研究者さんの登壇が多く、実際に何が必要とされているのか感じ取ることができてとても良いと思います。図書館関係者だけだと研究の現場で求められていることから外れた独りよがりになりそうで、視野を広く持つことの大切さが身にしみます。でも、研究者さんの考えていること、必要としていること、困っていることを聞くたびに図書館が考えていることと距離を感じ、どうしたものだろうかと途方にくれます。研究者は機関を超えて分野ごと

のグローバルなコミュニティでものごとを考えています。研究機関の図書館の役割はあるのでしょうか？必要とされる図書館になるには、どうしたらよいのでしょうか？〔動画中継視聴者〕

(その他/研究者)

・状況確認、update に役に立った。研究者・論文執筆者視点（プレプリントを知らないシニアな研究者・学協会長などが研究者に／学協会に／各自所属機関にとって役に立つ可能性を感じるのかどうかなど）からの議論を深めてもよいのではとも思いました。



## 企画後記

😊 企画のプロセスで不馴れだったため周りに迷惑をかけてしまいました。しかしながら、結果としてプレプリントに関してさまざまな分野の研究者のみならず、SSRN の Gordon さんにも運良くお話ししていただくことが叶い、大変刺激的なセミナーになって良かったです。

坊農 秀雅

(情報・システム研究機構

ライフサイエンス統合データベースセンター)

😊 今回 twitter 発信を担当しました。今回のテーマ“プレプリント”は日頃の業務ではあまり意識しておらず、大変勉強になりました。研究の速度が増していることがプレプリントを促進している、との言葉が印象的でした。各分野の研究者の具体的なお話を聞くことができ、プレプリントの分野ごとの浸透状況や出版社と研



究者の関係性など、常に最新の動向を認識することの重要性を痛感しました。

笹渕 洋子

(早稲田大学図書館)

😊 今回のセミナーは、プレプリントに重点を置いた企画でした。図書館職員の方にはまだ自分たちには馴染みが薄いと思われたのかもしれませんが、コーネル大学の図書館が arXiv.org を運営しているように、そう遠くない時期に図書館職員の仕事の一つとなる日が来るのかもしれませんが。オープンサイエンス時代の大学図書館の新たな機能のひとつとして、これからもプレプリントには注目していきたいと思います。

梶原 茂寿

(室蘭工業大学附属図書館)

## 第2回 SPARC Japan セミナー2017

「プレプリントとオープンアクセス」

### 開会挨拶/趣旨説明

坊農 秀雅

(情報・システム研究機構 ライフサイエンス統合データベースセンター)



#### 坊農 秀雅

理化学研究所においてFANTOM (Functional annotation of mouse) プロジェクトの立ち上げに関わった後、埼玉医科大学ゲノム医学研究センターを経て、2007年7月より情報・システム研究機構 ライフサイエンス統合データベースセンター (DBCLS) にて統合データベースプロジェクト (統合DB) に従事。統合DBの広報・普及活動として統合TVや統合データベース講習会AJACSの立ち上げに関わり、現在は国立遺伝学研究所にて日本DNAデータバンク (DDBJ) と連携して大規模塩基配列データ利用環境の構築とそれを利用した生命科学研究を行っている。京都大学博士 (理学)。



第2回 SPARC Japan セミナー2017「プレプリントとオープンアクセス」にご参加いただき、誠にありがとうございます。本日のセミナーでは、四つの講演の後、全体議論が行われます。プレプリントに関するセミナーはこれまであまりなかったので、議論して新しい知識を得られればと思います。

#### 注目される bioRxiv

私は1年前の9月に行われた第1回 SPARC Japan セミナー2016「オープンアクセスへの道」で、生命科学において、研究者はどのような雑誌に投稿するかという話をしました。図1はそのときに出したスライドの1枚です。私自身は生命科学研究者で、生命科学研究者から見たオープンアクセスはどのようなものかと考えたときに、基本的には「ゴールドオープンアクセス」と呼ばれる、オープンアクセスの費用を論文を出した投稿者が持つという選択肢ほぼ一択で、オープンアクセス＝ゴールドオープンアクセスという状況でした。

「グリーンオープンアクセス」と呼ばれる、機関アーカイブなどに出すタイプのオープンアクセスはその

言葉自体あまり知られていませんでした。その原因は、インパクトファクターと呼ばれる数値が付いてない雑誌には出したいくないという研究者が生命科学系には多いことです。

もう一つは、ピアレビューを経た論文のみを業績とするという伝統があるからだということを昨年9月にお話ししたのですが、それ以降、急速に情勢が変わって、生命科学の分野においても bioRxiv というものが昨年の後半ぐらいからいわれるようになり、前から bioRxiv 自体はあったのですが、無視できなくなって

#### 生命科学研究者の立場から見た Open Access (OA)

##### ・ 現状、ゴールド OA一択

- ・ すなわち「オープンアクセス」＝ゴールドOA

+ BioRxiv

##### ・ グリーンOAという言葉自体ほとんど知られてない

- ・ IFの付いてない雑誌は回避される傾向
- ・ ピアレビューを経た論文のみ業績とみなす伝統

第1回 SPARC Japan セミナー2016「オープンアクセスへの道」  
坊農秀雅「生命科学分野における研究者の投稿先雑誌選択傾向とOAへの意味づけ」  
スライド一部改変

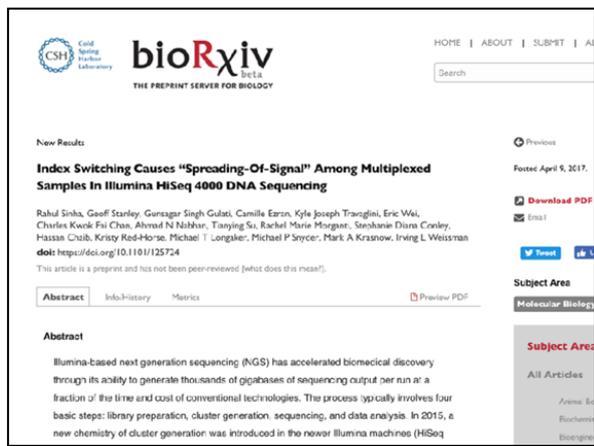
(図1)

きました。余談ですが、これは「バイオアーカイブ」と読みます。まだあまり広まっていないのですが、かなり注目されてきています。bioRxiv に挙げられた論文の中で、この1年で大きなインパクトがあったと思うのは、図2です。

最新の次世代 DNA シーケンサー「HiSeq 4000」に不具合がある、データがおかしいのではないかという論文が今年4月にピアレビューを経ずに bioRxiv にアップされ、われわれの業界では非常に問題になりました。実際にこのシーケンサーから出てきたデータの幾つかはおかしいのではないかと騒がれていて、現状でもこれはまだ解決していないのではないかと思うのですが、そういう状況が非常に素早く研究者に伝わったということで、このプレプリントは非常に注目されるようになってきて、それ以外でも最新の機械に対する評価は bioRxiv で発信されるということが起きています。

今回のセミナーの準備をしている間にも、「Science」誌に、プレプリントに関する記事が幾つか出ていました(図3)。エディトリアルにおけるプレプリントのエコシステム、プレプリントとどう付き合っていくか、「Science」はそれをどう考えるかというようなことが書かれているので、これも図書館に行って読んでみてはいかがかと思います。ここでは詳しくは述べません。

私自身の専門である生命科学分野の話ばかりでしたが、今日はそれに限らず、プレプリント全般に関してお話をいろいろな分野の方からしていただきます。



(図2)

プレプリント自体や arXiv というプレプリントサーバはそれこそ約30年前から進められています。生命科学分野で急に注目されているので、そこに出そうという人も出てきているということです。

それ以外の分野においても、例えば化学の分野においても ChemRxiv というプレプリントのサービスが公開されています(図4)。商業出版社においても、Elsevier が SSRN というリポジトリで参入しています。

### 本セミナーの趣旨

本日は、まずプレプリントサーバに関わってこられた引原先生にお話しいただきます。そして、変化の渦中にある、化学分野と生命科学分野の研究者お二人に、普段されている研究活動も含めてお話しいただき、その中でオープンアクセスの今後についても話し合いたいと思います。SSRN のマネージングディレクターで

**'Preprint' discussed in Science**  
**29 SEPT 2017 VOL 357, ISSUE 6358**



- Preprint ecosystems in Editorial (p1331)
- THE PREPRINT DILEMMA (p1344-1347)
  - Biologists are posting unreviewed papers in record numbers. Here's a survival guide
- How biologists pioneered preprints—with paper and postage (p1348)
  - Francis Crick and James Watson experimented with preprints as well as DNA.

(図3)

**趣旨説明: プレプリント**

- arXivは約30年前から
- 生命科学で急に注目(BioRxiv)
- プレプリントサービスの提供に参入
- 国際学会(ChemRxivのACS他)
- 商業出版者(SSRNのElsevier)






(図4)

ある Gregg Gordon さんにもお話を伺って、議論を進めます。

この会場の内容は YouTube live で配信されていますし、質問事項等は Twitter で発信していただけます。ハッシュタグ (#sparcjp201702) を付けていただくとわれわれも見つけやすく、こちらで質問にお答えすることもできると思いますので、ネットで見ている方は Twitter でハッシュタグを付けて、どの先生への質問かということも明記していただいて発信していただければと思います。

## 第 2 回 SPARC Japan セミナー2017

「プレプリントとオープンアクセス」

# arXiv.org の次世代システムの公開と戦略

引原 隆士

(京都大学図書館機構長/arXiv.org 会員コンソーシアム代表)

### 講演要旨



arXiv.org は、1991年にロスアラモス国立研究所で物理学分野のプレプリントサーバとして設立され、その後、数学、非線形科学、計算機科学、統計学、計量ファイナンス等へ領域を広げた。2011年に Cornell 大学図書館が運営している。本年、新たに電気システム科学、計量経済学のカテゴリーを増やし、次世代システムの概要を固めた。ボードメンバー会議から、新システムのアイデアと arXiv.org の今後の戦略について情報提供する。



### 引原 隆士

1987 京都大学大学院工学研究科電気工学専攻博士課程修了。2001年京都大学大学院工学研究科教授。京都大学工学博士。この間コーネル大学客員研究員。2012より 京都大学図書館機構・機構長・附属図書館長。現在、内閣府オープンサイエンス推進に関するフォローアップ検討会委員、文部科学省第9期学術情報委員会委員。

私はもともと arXiv.org (アーカイブドットオルグ) のユーザーであり、また、Member Advisory Board (MAB) という arXiv.org のアドバイザーボードに、国立情報学研究所から選出されて、昨年度からメンバーとして参加しています。

arXiv.org がカバーしている分野は主に数理物理ですが、最近はバイオロジーが加わったり、さらに新しい分野が追加されています。こういう分野について arXiv.org が何をしてきたかということをお話しすると同時に、今後どういうことを目指しているか、私の知る限りでお話ししたいと思います。YouTube に、arXiv.org がどういう歴史をたどってきたか、何をしてきたかが出ているので、参照していただければと思います (<https://www.youtube.com/watch?v=ntoxZzh0ha8>)。

### arXiv.org のこれまで

arXiv.org の創始者は Paul Ginsparg で、ロスアラモス国立研究所の彼のデスク上の HP のサーバから arXiv.org は始まっています。それが 1991 年で、インターネットが始まる前です。私が在外研究に行ったのが 1993 年で、その直前は BITNET などが連絡手段でした。その時代にメールベースで始められたのがそもそもその始まりです。

それをまとめると図 1 に書いてあるとおりです。高エネルギー物理学の論文のやりとりを創始者の Ginsparg が始めました。お互いに論文をやりとりして、共有して、必要なものをこのメールベースのデータベースに早く入れるようにする、あるいはそれをアーカイブして置いておくということが行われました。

それから 20 年たったことをきっかけに、2011 年にコーネル大学の図書館に移行しました。私自身が在外

研究に行っていたのがコーネル大学で、それは 1993 ~1994 年ぐらいでした。その当時はライブラリはあまりこういうことを意識していませんでした。昨年、コーネルライブラリのミーティングに久しぶりに行って、どういう形でここに移行したかという経緯などいろいろお聞きしました。それが今、エコシステムというか、そもそも始められたシステムをどうこれから発展させていくかという大きな契機になっていると思います。

分野については、1991 年に高エネルギー物理学の分野で始まっています。その後、1997年に数学、1998年にコンピュータサイエンス、2003年に quantitative biology、2007年に statistics、2008年に quantitative finance で始まり、ここでいったん発展は止まっています。ところが、今年 2017 年 9 月 18 日に EESS (electrical engineering system science) が加わりました。

### Short history of arXiv.org

- 1991 GINSPARG, Paul, **Repository Alert System**  
hep-th@xxx.lanl.gov (before internet)  
High Energy Physics
- 2011 → Cornell University Library
- Categories Expansion:  
Physics (1991),  
Mathematics (1997),  
Computer Science (1998),  
Quantitative Biology (2003),  
Statistics (2007),  
Quantitative Finance (2008),

---

EESS (Sep. 18, 2017), Econ (Sep. 26, 2017)

GINSPARG, Paul. "ArXiv at 20". nature. vol. 476, p. 145-147, 2011

(図 1)

### DIGITAL PIONEERS LEAD THE WAY TO SHARING RESEARCH ONLINE

The popularity of the arXiv preprint server has grown remarkably since its launch in the early 1990s. Academics enjoy the universal access, low cost and speed of online distribution.

**PHYSICS ENVY**  
Mathematicians, astrophysicists and even some biologists have joined high-energy physicists in uploading articles to ArXiv.

150,123 High-energy physics	125,157 Condensed matter	123,902 Astrophysics	125,689 Math	28,326 Computer science	25,443 General relativity	22,572 Nuclear physics	35,154 Quantum optics	35,156 Vibrations
--------------------------------	-----------------------------	-------------------------	-----------------	----------------------------	------------------------------	---------------------------	--------------------------	----------------------

694,000

Total uploads to ArXiv (as of 8 August 2013)

GINSPARG, Paul. "ArXiv at 20". nature. vol. 476, p. 145-147, 2011

(図 2)

さらに、economics が 1 週間後の 26 日に加わっています。

物理で始まって、物理の関連で次に数学が始まったというのは分かりやすいのですが、この後どのように発展していくかということを決めて、現在に至っているかというのが重要なキーになります。

図 2 は 2011 年に 20 周年を記念してまとめられたものです。20 周年の時点では、月ごとのサブミッションの増加は左上のグラフのようになっています。週ごとのダウンロードの増加は右上のグラフのようになっています。指数関数的に増えているということが見て取れると思います。

このダウンロードの対象あるいはサブミッションの対象になっているのが下のマップに示されている分野です。このような比率で入っていたのが 2011 年です。

今後これがどう変わっていくかというのが arXiv.org の中では非常に重要な観点です。これから広がっていく分野を加えつつ、それまでの分野を arXiv.org として維持していく、その流れをどうやって判断していくかというのが大きな課題になっています。

プレプリントサーバとしては、arXiv.org は 1991 年に始まりましたが、bioRxiv は 2013 年に始まりました。それ以外にも図 3 のようなものがあります。arXiv.org と bioRxiv は今、かなり連携しています。お互いの運用に関して情報交換しており、これが今の bioRxiv の発展をサポートしていると思っていただければいいと思います。

### Preprint archive families

arXiv.org	1991	1.27M
bioRxiv	2013	14K
engrxiv	2016	130
SOC ARXIV	2016	1,400
BioRN	2017	4K
PsyArXiv	2016	800
LawArXiv	2017	300
AgriXiv	2017	12
ChemRxiv	TBC	0

← Publishers' archive

**Their archives are business.**

(図 3)

ソーシャルサイエンスの SSRN など、パブリッシャー側に移ったアーカイブもあります。パブリッシャー側に移ったアーカイブは、もうビジネスに移ったものと見ています。ですから、研究者側にあるアーカイブ、ビジネス側にあるアーカイブの二通りあるということです。これを同一に話ができるかという非常に難しい面があります。

arXiv.org は昨年 2016 年からドネーションを始めています。ドネーションのページがあり、ダウンロードする側とサブミッションする側、両方のユーザーに、このシステムを維持するためのドネーションをお願いしております。これがかなり大きな額で、arXiv.org を支えています。

### 統計で見る arXiv.org

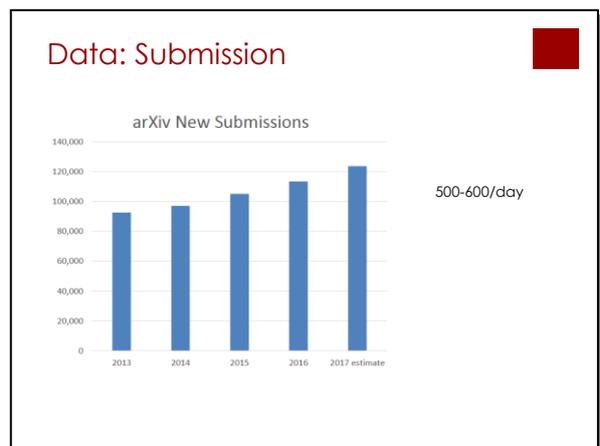
arXiv.org のサブミッション数は 2012 年に約 84,000 で、それからどんどん増えていって、2013 年に約 92,500、2016 年に約 113,380 です。ダウンロード数は 2016 年に 139 百万です (図 4)。ですから、膨大な数のダウンロードが始まっているのです。

中にどれだけ面白いものがあるかということが当然重要ですが、arXiv.org の趣旨は、パブリッシュされる前の論文ができるだけ早い段階でユーザーに伝えられるということです。今はこれをいろいろな検索エンジンがウォッチしていて、いろいろなところからこの情報が上がったということを出せるような仕組みに変わりつつあります。ですから、当初のメールベースのも

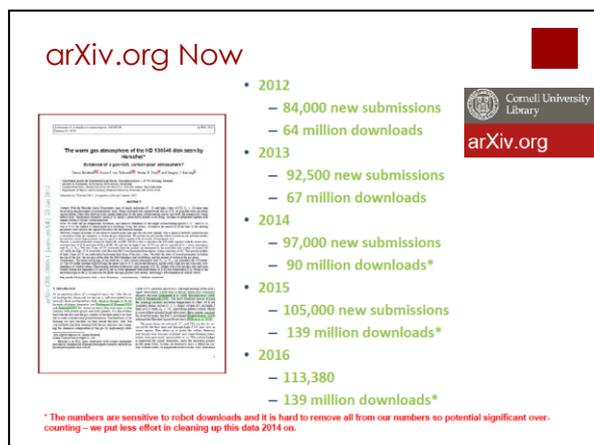
の、それからインターネットが始まったときにこのサイトにアクセスしたら見られたというものから、検索エンジンを介して提供するという状態にだんだん移っています。

図 5 はサブミッションの変化です。2017 年現在で約 12 万あります。1 日に 500~600 サブミッションされます。これを処理しなければいけないのです。この処理をコーネル大学のグループが今行っています。

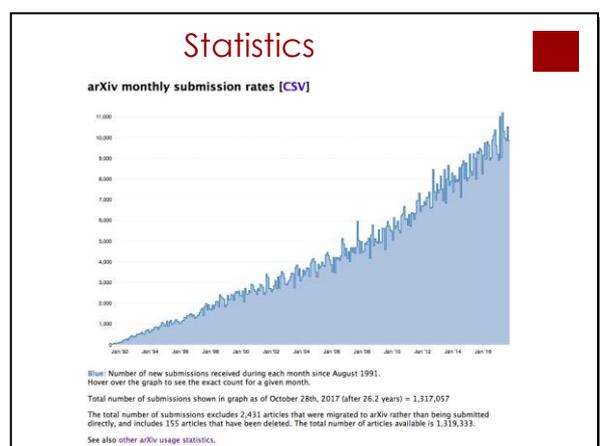
それをもう少し統計的に見ると図 6 のようになっています。1992 年は微々たるものだったのですが、現在はここまであります。サブミッションは、日本とは違い、夏休みはやはり少なく、クリスマスも少ないです。この波が大体毎年同じような傾向をたどっています。これは増加を総量で書いているので、その年々で出すと大体フラットな変化をしています。加算してリニアに増えています。



(図 5)



(図 4)



(図 6)

これではよく分かりませんので、もう少し分野ごとに見てみます(図7)。左はトータルの数で見たものですが、割合に変えると右のようになります。当初は高エネルギー物理学だったので青色のところほとんどだったのですが、それが時がたつにつれて大きく変わっていきます。Condensed matter physics が増えている時期もあります。Astrophysics、condensed matter physics が 2000 年あたりに高エネルギー物理学と同じぐらいの比率になります。その時期に数学が少しずつ増えていき、現在では高エネルギー物理よりも数学の割合が高いです。総数が増えているから、総数の中の数で見ていただかなければいけないのですが、他の物理やコンピュータサイエンスも増えていることが分かります。

この変化を見ながらサーバとしての運営をしていて、例えば mathematics の中でも細かいカテゴリーを付けていく、あるいは condensed matter physics の中でも細かいカテゴリーを付けていくことによって見えやすいものに変えていくというような変化をさせています。

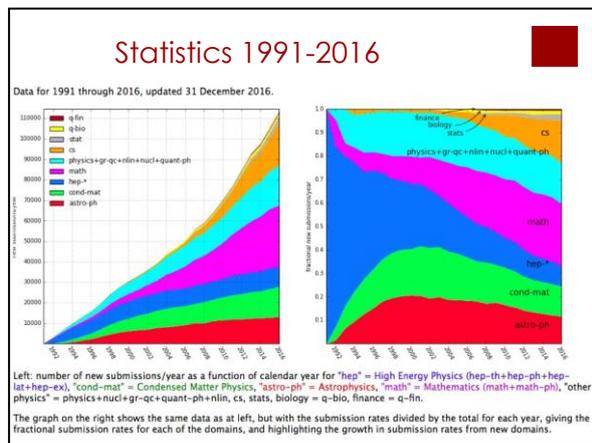
このうち、サブミッションとダウンロードを行っている国の内訳は図8です。まず、arXiv.org を支えているアメリカのコンソーシアムが 18%、アメリカのノンコンソーシアムが 21%と結構多いです。一つ一つの大学や研究所がコンソーシアムとして組んでいる部分と、そうでない部分が大体同じ比率になっています。UK が 7%、日本が 7%と非常に多いです。この中でも研究大学といわれるところが多く論文をサブミッ

ション・ダウンロードしています。スウェーデン、オランダも結構数が多いです。ドイツが面白くて、物理系で投稿するときには自動的に arXiv.org に上がるようなシステムを組みつつあるので、だんだん割合が増えているというのが現状です。

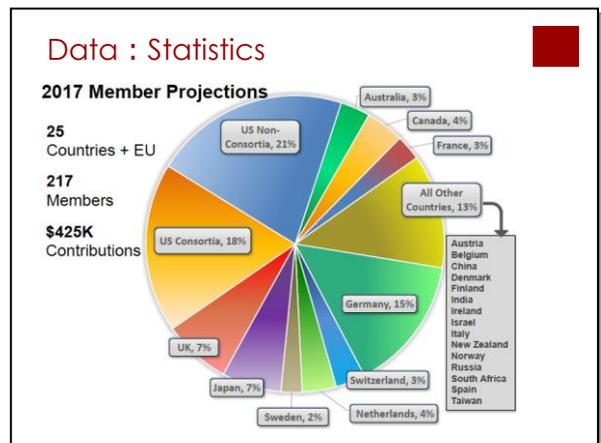
アジアでも増えており、中国のユーザーの割合が非常に高くなっています。また、投稿よりもダウンロードの数が非常に多いということが知られています。このような割合になっています。

これを幾つかのサーバで支えています。以前は日本にもサーバがあり、京都大学の基礎物理学研究所にミラーサーバを置いていたのですが、それは約 2 年前(2015 年 12 月)廃止になって、現在は 14 のサーバがあり、そのうち五つのミラーサイトが中国、ドイツ、インド、スペイン、ロスアラモス国立研究所にあります。これらの国からアクセスするのが非常に難しい場合があるために置いています。日本の場合、ネットの回線が非常に強くなっているので、あまり問題なくなったのですが、中国やインドなどはミラーサーバを置くことによってアクセスがスムーズにいくようになりました。ドイツに関しては東ヨーロッパ系のところをサポートしているような形でしょうか。

もとのサーバシステムもまだコーネル大学の中にあります。以上のシステムで今は運用していますが、これをそれぞれの国でスタンドアロンでマシンの上に置くだけではなく、今は Amazon のウェブサーバのクラウドに移しつつあります。Amazon に移してしまう



(図7)



(図8)

ことがいいかどうかというのは議論がありますが、移して、よりアクセスがしやすいものにする方向で動いています。

問題は、サーバはいいのですが、このデータベースの部分とどうやってリンクを維持していくかということです。先ほど、データが非常に増えているということをお見せしましたが、そのデータベースに全部リンクしないとイケないのです。それをやるのは、マシン間では非常に難しいということが明らかです。

ですから、今、クラウドの上に持っていこうとしているのですが、そのときにやはり Amazon だけでいいのかという議論はされています。現状は様子を見るということで、Amazon のクラウドに移すという方向で進めていますが、今後もこれがずっと続くかどうかというのはまだ分かりません。

次は、どこから論文にアクセスしているかです(図9)。最初はサーバに直接アクセスして、そのサーチを使って論文を見るという形だったのですが、2年ほど前のデータでは、本体の arXiv.org のホームページからのサーチ以外に、Google のサーチエンジンを使っているもの、Google Scholar を使っているもの、ADS を使っているものがあります。それから arXiv email alerts とありますが、そもそも、arXiv.org は論文をアップしたときに、アップされたというカテゴリーごとの e メールをアラートで送ってくるシステムだったので、そのシステムがまだ生きています。今、他のところでも同じようなシステムを使っています。「あなた

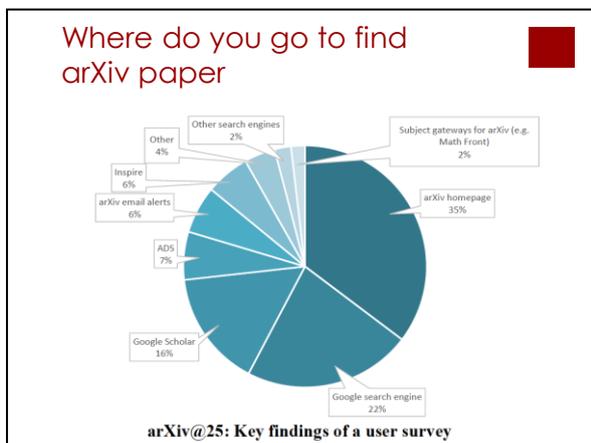
の論文が引用されました」というようなアラームがあると思うのですが、それが結構有効だということがよく分かります。

INSPIRE は別のデータベースサーチになりますが、こちらのリンクも結構強くて、CERN などからこちらを経由しています。これは重要なデータで、どこにどのように新しいデータベース、新しい arXiv.org が、システムを組んでリンクを付けていくとサーチが進むかというのが分かりやすいデータになるかと思います。これが状況です。

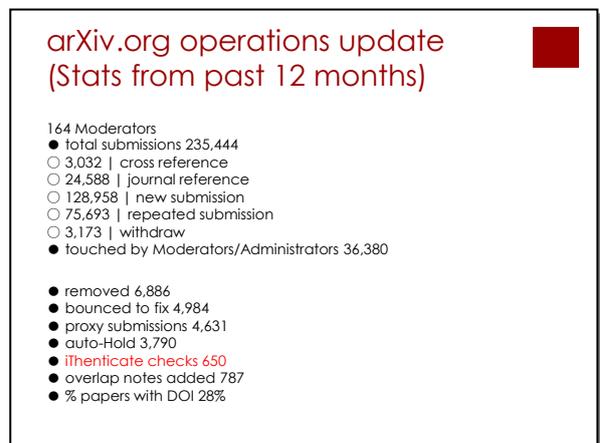
少し中身に入ります。先ほどトータルサブミッションの話をしました。そのトータルのサブミッションは約 230,000 あるのですが、そのうち、ジャーナルのリファレンスが約 25,000、クロスリファレンスが約 3,000、ニューサブミッションが約 130,000、再投稿が約 75,000、withdraw (取りやめ) が幾つかあります(図10)。

これらをアドミニストレーターが処理するのですが、非常に大きな数を扱っていることが分かるかと思えます。こういうことをどのような人たちがやっているかは、オープンになっている情報からご覧いただければと思います。

この中でリムーブされたものが約 6,800、それから、特に気を付けて見ていただければいいかと思いますが、iThenticate、要するにチェッカーで約 650 がリジェクトされています。だから arXiv.org のユーザーも決して紳士ではないということがよく分かります。



(図 9)



(図 10)

iThenticate で引っ掛かるというのはよほどのものだと思いますが、こういうものが出てくるということです。これには説明のとき、みんな苦笑いしていました。

それから、DOI を付したものが約 28%あります。このプレプリントでも DOI や ORCID などを著者のメリットのために使うという考え方をしています。

### arXiv.org の目指す方向

研究をエンカレッジするというのが arXiv.org の基本的な考え方です (図 11)。研究してください、論文を書いてください、投稿してくださいとエンカレッジしているわけです。そのときにリファレンスのペーパーを見えやすくして、アーカイブもきちんとしてあげましょうというのが arXiv.org の考え方なのです。

一方、arXiv.org が何をメリットとして主張するかというと、当たり前のことですが、ページチャージは要りません。それから、論文を投稿して掲載されるまでに何年かかからず、1~2 週間でプレプリントが上がるということです。化学系や物性系だとすぐに掲載されることが結構あり、何年かというものが皆さんにあまり伝わらないケースが多いのですが、システム系や数学系、物理でも理論系だと、1 年たっても 2 年たってもまだ査読中ということが結構あります。私も若いころに 1~2 年かかるということがあり、このままいくと私は首になるというぐらい、たくさん出しているのですが、査読が終わったものが何も出てこないのです。どんな論文を出しているのかも、著作権の関係で出せ

ないのです。

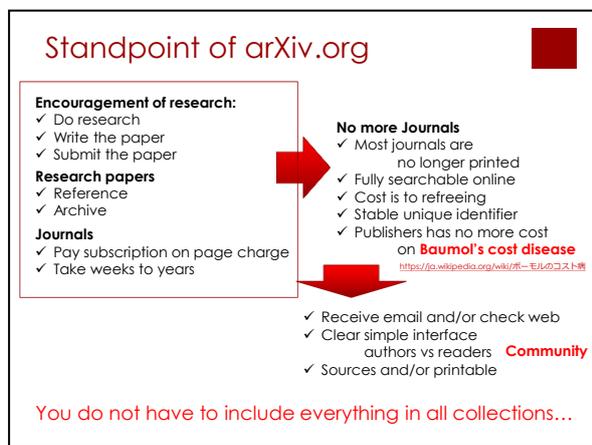
そういうものに対して、出したということをしきりと arXiv.org でオープンにして、それが伝わるようにして、物理なら物理の世界が早く進むようにエンカレッジしていくというのが基本的な考え方です。ですから、やはり物理特有の状況に根差してこうなってきたというのがよく分かると思います。

こういう立場で始まったのですが、現在、運営者たちはこう言います。「もう論文は要らない。どんな論文ももうプリントされていないではないか、ほとんどがオンラインでサーチできるようになっている、コストはレフェリーのお金にすぎない」と。レフェリーは研究者がしています。よく日本でも産地直送と仲買を経た場合の値段の違いという話がありますが、産地直送でいいのではないかという話になるわけです。

パブリッシャーは、「ポーモルのコスト病」といって、もはやもうける仕組みにはなっていません。「ポーモルのコスト病」というのは Wikipedia を引いていただいたら分かると思うのですが、増やすことによってもうかるシステムではない、本来は価値で動いているのにそれを数で処理している、運営者たちは、そういうシステムの中に入る必要はないということを言っているのです。

ではどうしていくかということ、e メールを使うか、ウェブを使うかは別として、コミュニティをもっとクリアにしていく、コミュニティ間のやりとりをエンカレッジしていくような方向にして、著者とリーダーが直接やりとりできるシステムとして維持したいということです。それから、ソースはプリンタブルにしておくということができれば、自分たちが好きなように使えるということになります。

多くの場合、システムが大きくなってくると、もっと入れてください、この分野も増やしてくださいなどいろいろ言われると思うのですが、全ての分野を入れるつもりはなく、自分たちの考え方で維持できる分野だけを増やしていこうとしています。そういう共通認識を持っている分野を探し出して動かしているという



(図 11)



システムがあって、サブミッションがあったときに、ビューワー側とアドミニストレーターが見るサブミッション側の処理と、モデレーターがお互いを見ながら全体のシステムを運営していくという流れをつくっています。

ビューワーについては、今までは図 16 のようなビューワーで使いにくさがあるということを皆さん分かっていて、このビューワーに対して他の方法を示しています。

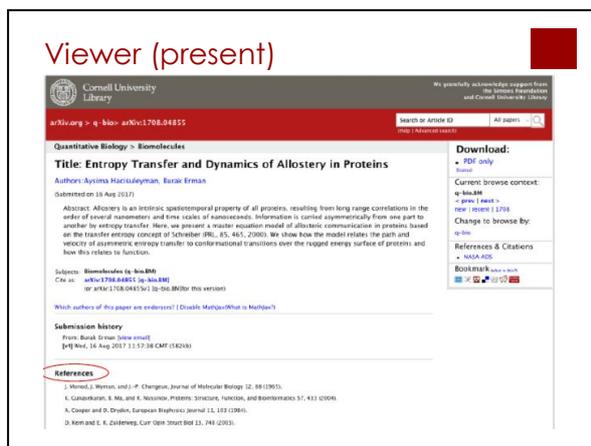
例えばリファレンスだけを別で見られるようなものです(図 17)。こういうのはよくあります。このように変えたのですが、アンケートを取った結果、前の方がいいという人が結構多いのです。アンケートの結果、49%が前の方がいいと答え、51%が新しい方がいいと答えました。要するに、ビューワーは慣れたものがいいということがよく分かります。

システムの運用は図 18 のようになっています。リーダーグループがあって、私が属している Member Advisory Board (MAB) というアドバイザーボードがあって、Scientific Advisory Board (SAB) が開発をしています。それから、コーネル大学のライブラリのアドミニストレーションがサポートしています。

これを非常に少ない人員で行っています。図 19 の資料もオープンなものです。full-time equivalent (FTE) ですから、1日を1とすれば、0.1~1のエフォートで、エフォート管理をしながら全体を運営しています。ライブラリとこちらの仕事を共有しながらやっているというのが現状です。

### パブリッシャーとの関係

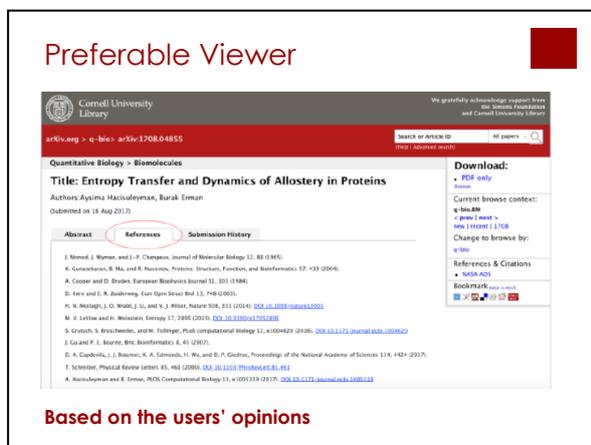
ここが今日一番重要かと思いますが、パブリッシャーというものをどう取り込むか、どう連携するかとい



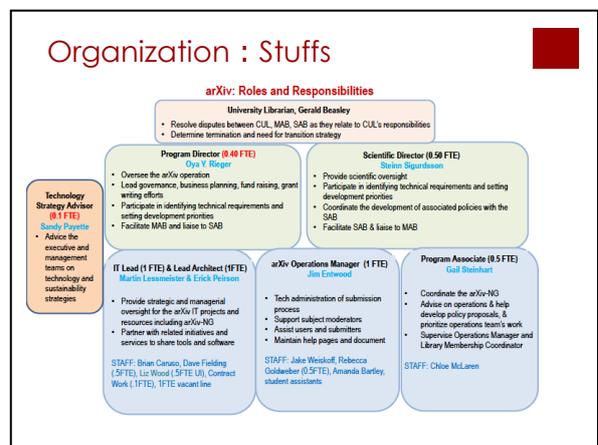
(図 16)



(図 18)



(図 17)



(図 19)



TeX システムもバージョンアップしています。arXiv.org は PDF で出すこともできますが、TeX のソースからアップすることも当然できます。物理系は TeX でやっているケースが多いので、それがそのまま使われています。

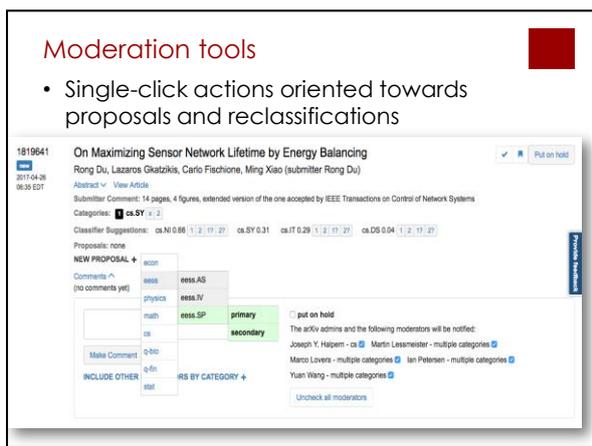
図 25 はシステムのハイライトです。いろいろなソフトを使っています。ネクストジェネレーションのアーキテクチャは、まとめると図 26 の 5 項目になります。インテグレーションが非常に重要なのですが、その前のテクノロジーを現代風書き換えるということをやっているのです。それによってセキュリティの問題が当然起きてきます。セキュリティの問題に関しては今もやりとりをしていて、チェックしているという状況です。ですから、完全に確保できたという状況ではないと思っています。

図 27 は、全体の運営としてどういうサステナビリティ

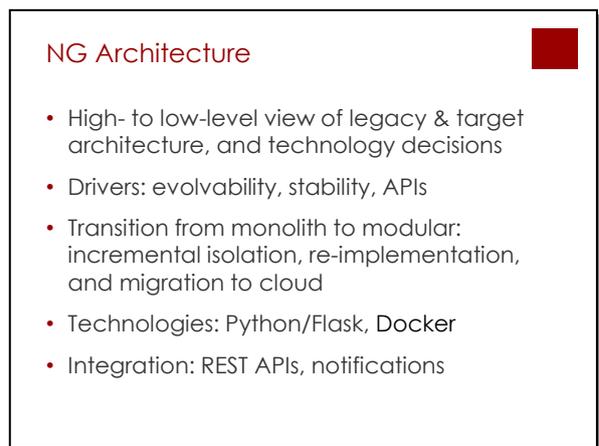
ティをどういう考え方でやればいいのかというマップです。これは Oya Rieger さんが全体の運営の中でまとめたものです。バランス良く全体を運営したいと主張しています。そのソースになる、ファンドによるお金、各コンソーシアムが支払うお金、大学が支払うお金をどう当てはめていかないといけないかが描かれています。本当は予算の関係をお見せしようと思ったのですが、ビデオで流れるので今日は割愛させていただきました。ウェブで見られますので自分で探してみてください。

**結論**

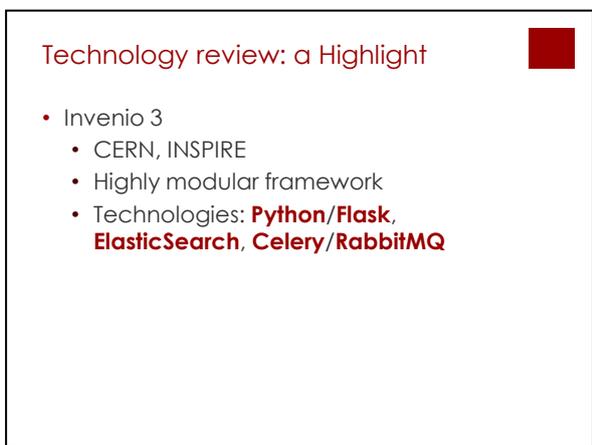
オープンライセンスに対するライセンスは何かという議論がずっと続いています。また、エンバージした後、ハイブリッドで出しているものと、ももとのパブリッシュされたものとの関係がまだ微妙なところが



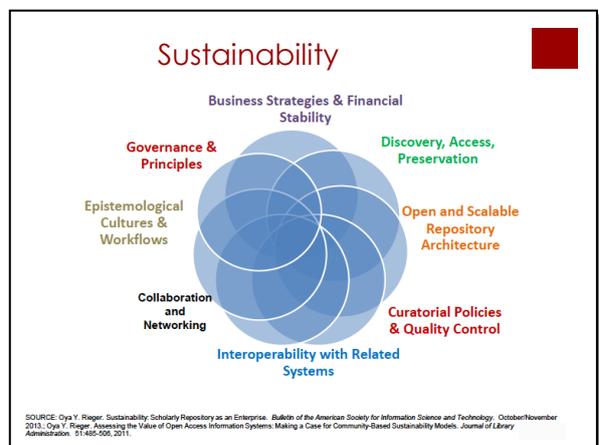
(図 24)



(図 26)



(図 25)



(図 27)

あります。エンバゴしているからいいのではないかと  
思われますが、何が二次的という話になったとき  
に、どちらがソースかというのが分からない面あり  
ます。そういう部分が残っています。

arXiv.org の話の中でこれは共通認識だと思うので  
すが、別にパブリッシングモデルを変えたいと思っ  
ているわけではないのです。ユーザーを変えたいの  
です。ユーザーが変わってくればシステムはもっと  
発展するし、今のパブリッシャーから多く要求され  
ている支払いを抑えることもできるはずですが、  
パブリッシャー相手に訴訟問題を起こしても仕方  
がないので、ユーザーが自発的に変わっていくよ  
うなシステム、それがやりやすいようなシステ  
ムをつくり上げていくというのが今からの流れ  
です。ネクストジェネレーションのシステムも  
これが変わるものではありません。ですから、  
これからパブリッシャーとして arXiv.org が  
出ていくというわけではなく、やはりコミュニ  
ティがどうやって自分たちを守るかという考  
え方に根差していると思っただけならばよろ  
しいかと思えます。

機関リポジトリや他のアーカイブと協調して  
いくことが必要だというのは認識されています  
が、例えばリポジトリ関係は、日本の考え方  
とアメリカの考え方、ヨーロッパの考え方  
が全く異なります。日本のリポジトリはど  
ちらかというプレプリント的なものになっ  
ているので、既にアーカイブになっているの  
ですが、アメリカの場合はそうではなくて  
ビデオや講義録などが中心です。ですから、  
一括して同じようにはできないというこ  
とが言えます。

逆に言えば、日本の場合、リポジトリは  
既にプレプリントの機能を持ちつつあるとい  
うことです。ですから、それを発展させるこ  
とによって、うまい連携ができる可能性は  
あると思います。そういう議論も今年提案  
させていただいたのですが、やはり国情によ  
って考え方が違うので、プレプリントサー  
バとリポジトリの関係は日本なりのもの  
を構築していく必要があるだろうとい  
うのが結論です。

●フロア 1 国立情報学研究所の武田です。今後、シ  
ステムをサステナブルに維持するということに  
関して、資金はファウンデーションの資金、  
コーネル大学の支援、コミュニティ（事実上  
は各大学等）の支援から成っており、一方  
で 2017 年も分野を広げることについて  
ですが、コストと、分野を広げることにつ  
いての議論はあるのでしょうか。

●引原 非常に微妙な質問なのですが、財団  
からの資金は主には次世代システムの開発  
に回っていると思っただけならばよいと思  
います。ですから、今年は昨年から予算が  
倍増しています。新しいシステムと旧来の  
システムを並列して走らせなければいけ  
ないということがあって、そのようになっ  
ています。

今はシステムを維持するランニングコスト  
は利用者が支払っています。日本でも NII  
をベースにコンソーシアムが支払いをし  
ているのですが、今、払っているところが  
利用者全体の 80%程度にとどまってい  
ます。残りは、これから増えてくれば  
当然徴収すると思えます。

それ以外のやり方は、先ほどミラーサイ  
トについてお話ししましたが、そういう  
運営で頑張っていたかということです。  
日本も昔そうだったと思うのですが、  
そういう形を取って、何らかの貢献を  
していただくような考え方になるか  
と思えます。

最後に、分野を広げるという話につ  
いては、分野はなかなか広がらな  
かったのです。今はほとんど 10 年  
ぶりに広がったわけですが、これは  
旧来のカテゴリの中で違う分野が  
起きてきているからです。それを  
外出したと思っただけならばいい  
かと思えます。EESS というのは「EE」  
と入っていますが、ほとんどは  
システムサイエンスなので、バイ  
オロジー関係でもシステムサイ  
エンス的なものは結構あります  
し、そういうものが今までのカ  
テゴリの中ではユーザーにリン  
クしにくい部分があったので  
外出したということです。外  
出したと同時に、もう少しそれ  
に関連する分野を増やしてい  
っています。エコノミクスも  
同じよ

うに、旧来あった分野を外出ししている、そういう形で少しずつ広げていますので、ドラスティックに変わったわけではないと思います。

●フロア 2 NICT の研究者です。私はこの分野に素人で、学会でプレプリントサーバの話を考えないかと言われて少し興味を持ちはじめたのですが、実は地球物理学でアメリカ地球物理学連合 (AGU) がプレプリントサーバを今立ち上げているのです。arXiv.org と少し微妙な関係のようなのですが、日本の日本地球惑星科学連合に対して一緒にプレプリントサーバをやりませんかという声を掛けてきました。日本側は、地球物理学はプレプリントサーバはなじみがなくて、私もそうなのです。

先取権やジャーナルで受理するかなど、いろいろな問題があるかと思うのですが、そういうのを学会でオーソライズして、われわれは先取権をプレプリントサーバ、ペーパーに認めますとか、われわれの発行するジャーナルは必ずプレプリントからでも投稿を認めますといったことを学会長自ら全員が承諾して、そのコミュニティがプレプリントサーバを受け入れないと、うまくいかないのではないかとこの雰囲気もあるので。実際に今動いているプレプリントサーバの上で投稿されている方々の分野では、コミュニティの規範、norm、カルチャーが既に確立していて、業績をそこで出していくようなプロセスが動きはじめていますか。

●引原 それもまた微妙なご質問だと思うのです。やはり分野によっても大きく違います。高エネルギー物理学では、SCOAP<sup>3</sup> の話もありますが、共有化、早く出す、グリーンオープンアクセスというのが文化として定着しているのだと思います。それをコミュニティが支えているという現状があります。

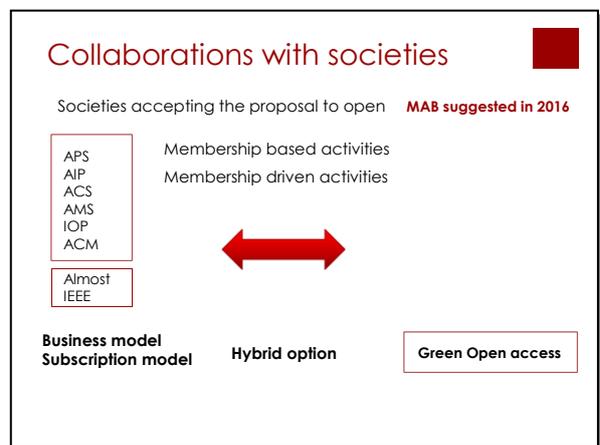
一方、先取権ということを非常にセンシティブに捉えている分野と、2~3 年かかるのだったら早くコミュニティに帰する方がいいと考えている分野と、ま

ちまちまです。ですから、物理の分野の考え方がどこまで他に適用できるかというのは、色合いを見定めなければいけないだろうと思っています。

新興のプレプリントサーバの危ないところは、ある程度データがたまってくると出版社が目付けます。そのときに規模の問題があって、小さいところは簡単に扱えます。M&A と一緒です。arXiv.org ぐらいになると、もう少し扱いきれない部分があります。クラウドになったら分かりませんが、そういう部分でやはり定着したというところの考え方を維持できれば、まだ大丈夫だろうと思っています。

コミュニティというのは、学会が残っている国は学会が持っています。そういうところは、出版を持っていても、基本的にプレプリントで出して先取権をらせて、あとは学会が最後のピアレビュー後の出版を受け持つのだというお互いの信頼関係を築くことが重要だろうということ、先ほどお見せした 6 学会 (図 20) と議論してアプローチしているという状況だと思います。

●フロア 3 NISTEP の林 (SPARC Japan 運営委員) です。クロスチェックでリジェクトしているとおっしゃっていましたが、もしご存じだったらどこで切っているのか教えていただけますか。あれはどこで切るのか実際に使ってみるとものすごく悩ましいのです。単純にパーセンテージが何パーセントだからという基準ではなさそうですが、あまり手間を掛けているようにも思



(図 20)

えず、ご存じの範囲で教えていただければ。

●引原 知っている範囲だけですが、それも質問がありました。実際にやって、これはどうだろうかとなったときに、エディトリアルとは言いませんが、少し上のレベルで判断するというプロセスを経ています。それで、機械的にやるのではなく、本人に問い合わせるということもしています。現に、私自身が問い合わせられたことがあります。そこはセンシティブなので、まめにやっています。

●フロア 3 結構、丁寧にやっているということですね。

●引原 丁寧にやっていると思います。

●フロア 4 日本学術振興会の関係者です。私は物理系なので比較的 arXiv.org には慣れているのですが、arXiv.org はプレプリントサーバですから当然プレプリントなわけですね。それが最終的に投稿したジャーナルの何巻の何ページに出たという情報が arXiv.org にどのくらいのパーセンテージで反映されているかはご存じでしょうか。それは本人がやらないとできないようになっているのでしょうか。

●引原 新しいものはよく分からないのですが、古いものに関して言えば、本人がその情報を書いた場合に示されています。メタデータがここ2年ぐらいで大きく変更されました。以前はファーストオーサーだけがメタデータに入っていて、セカンド、サード、それ以降は出ていなかったのですが、それが今は変更されています。

そういう段階になったので、著者管理に関しても非常に明確になってきました。ですから今後は、resubmission も含めて、きちんとどこに出たかというものも合わせて出す方向で進めるのだらうという議論にはなっていますが、まだ確定ではないです。

●フロア 4 ドイツでは自動的に arXiv.org に上がるシステムになっていると初めてお聞きして、何かそういう方向に進むのではないかという気がしています。

●引原 それはエンカレッジしています。例えば、日本の学会はどうですかと聞かれて、科学技術振興機構などをお願いするのがいいのではないかという話は当然あるのですが、オープンアクセスの考え方と整合を取らないといけないので、国としてきちんと定めた流れの中でやるというのが重要だと思います。

ドイツの場合はまさにそれをやっていて、現在、Elsevier とオープンアクセスに関して闘っていますが、その状況の中でこれは重要ですし、これと同時にやりとりするための ILL を構築していくというのがドイツの考え方です。入れる側と出す側の両側をきちんとグリーンにしていこうというのが戦略としてあるのだと思います。そこをサポートできると思います。

## 第 2 回 SPARC Japan セミナー2017

「プレプリントとオープンアクセス」

# 学術情報共有とオープンアクセスの未来

Gregg Gordon

(Managing Director, SSRN)

### 講演要旨



オープンアクセスとデジタルリポジトリは研究の相互交流において重要な要素である。これまでのグリーン OA とゴールド OA に最近加わったブラック OA はシンプルに回答できない課題を呈示した。『共有』はこの 20 年以上に渡って明らかに変わってきた。演者は SSRN の概要を説明し、社会科学分野のみでスタートしてからその範囲が生命自然科学を含め 30 を超える分野へ広がってきたことを紹介する。本講演では現実に起きている出来事や、広範囲での学術情報共有が引き起こす問題、オープンアクセスと学術情報流通の未来への考えを述べる。



### Gregg Gordon

KPMGやテクノロジーやヘルスケア関連の企業に勤務後、1994年にSSRNを創始したMichael C. Jensenを補佐する。学術研究と、イノベティブな研究をより早めるために必要となる変化について世界中で講演し、定期的に執筆している。最近の著書に、PLOS Biologyから出版されたThe Question of Data Integrity in Article-Level Metrics (共著)がある。

これから、Elsevier の土台であるオープンサイエンスと SSRN について、そして私たちがどんなふうに見えるかについて、お話しします。幾つかポイントを取り上げて、全体像を示したいと思います。

初めに、広い観点から Elsevier を見てみましょう。Elsevier は RELX グループの一部で、RELX はさまざまな企業を抱えています。どの企業も、情報の共有をベースにしています。RELX は、世界 24 カ国に 66 以上のオフィスを持つグローバル企業で、「The Lancet」「Cell」などの定評あるブランドを抱えています。出版業の変化に伴い、Elsevier も変化しています。オープンサイエンスに焦点を絞り、単なる出版社を超えたデータアナリティクス企業になりつつあります。

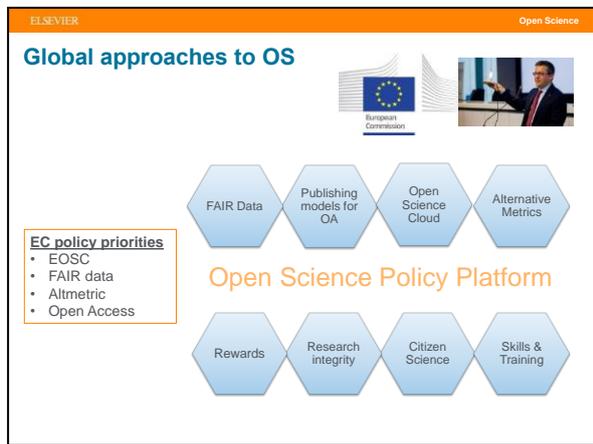
### オープンサイエンス

オープンサイエンスとは、研究をもっとオープンに、もっと協力的に、もっと透明性のあるものにするために協力して作業する方法のことで。ご覧のスライドには、「オープン」、「協力」、「透明性」という異なるポイントに貢献する Elsevier の幾つかの要素が示され

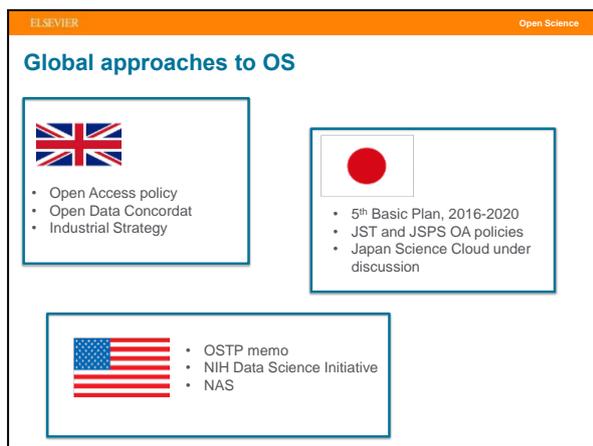


(図 1)

ています (図 1)。SPARC や ORCID など、世界中の組織が関わって、世界中でたくさんの活動が行われています。また、イギリス、日本、アメリカで、オープンサイエンスに向かう世界的なアプローチが取られています (図 2、図 3)。



(図 2)

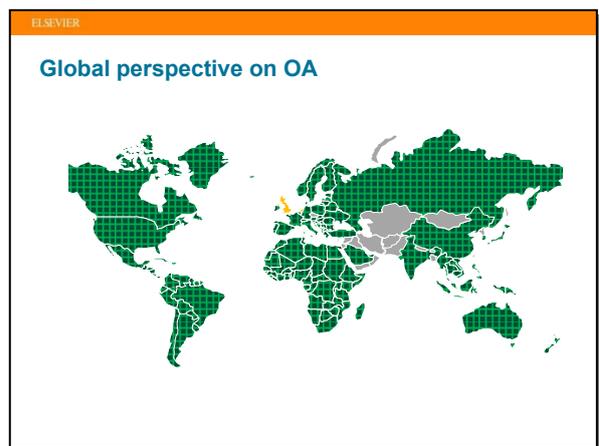


(図 3)

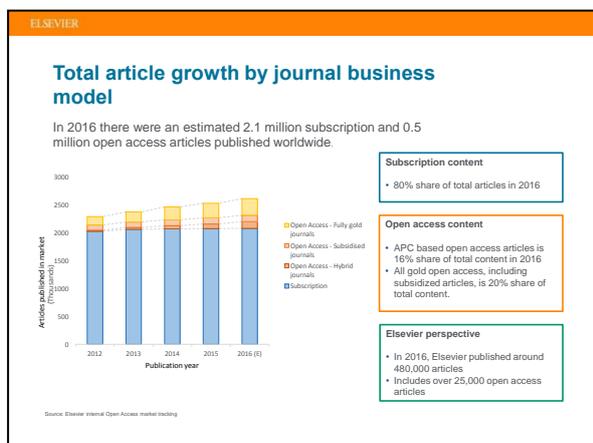
## よりオープンに

まずはよりオープンになることについて話しましょう。Elsevier におけるオープンアクセスの定義はシンプルで、誰でも学術研究に無料で永続的にアクセスできるということです。このオープンアクセスには、簡単に言えばゴールドオープンアクセスとグリーンオープンアクセスがあります。このグラフは、いろいろなタイプ別のオープンアクセスの伸びを示しています (図 4)。しかしながら、論文掲載の大半は依然として購読ベースです。この地図は、世界でのグリーンオープンアクセスの利用状況を示したものです (図 5)。これらの国々の多くにはゴールドオープンアクセスもありますが、グリーンオープンアクセスの方が世界中で利用でき、普及しています。

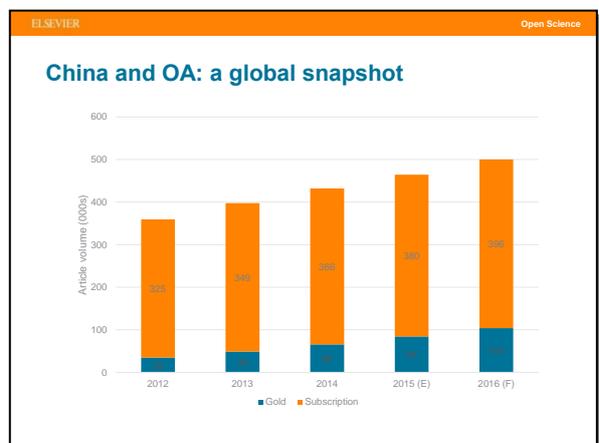
中国を例に見てみましょう。中国はアジア諸国の中で最も成長の著しい国です (図 6)。しかしながら、



(図 5)



(図 4)



(図 6)

成長は依然として限定的です。私は、これは私たち全員にとって、より共有を進めるチャンスだと捉えています。図 7 はゴールドとグリーンそれぞれの、Elsevier のオープンアクセスに関するデータの一部です。

もっとシェアできるようにするための方法の一つは、Elsevier のアプリケーション・プログラミング・インターフェース (API) を活用して、メタデータとコンテンツのいずれも機関リポジトリに入れることです (図 8)。Elsevier は CHORUS にも取り組んでいます。現在アメリカの諸機関と実験をしていて、日本とオーストラリアでも試験サービスを行っていますが、もっと幅広い機関リポジトリにコンテンツを提供するというものです。もっとシェアするためのもう一つの方法は、データの共有を始めることです。データはたいてい複数の研究者が利用しますから、シェアすればするほど、研

究者にとってメリットが大きくなります (図 9)。

データを論文にリンクさせると、特に特定分野のデータリポジトリに保管されている場合、とても役に立ちます。現在では、SSRN 内でデータを共有したり、多数の分野でデータ論文を発表したりすることが可能です。最新のパイロット試験ではオープンピアレビューの共有を開始しました。一般的なピアレビューの問題の一つは、査読者が匿名で、その功績が認知されていないことです。オープンにして査読者の認知度を上げることによって、もっと多くの査読者が査読をするようになると私たちは確信しています (図 10)。

### より協力的に

協力についてお話ししましょう。私は中国科学院の人と話をしたのですが、中国科学院はたくさんの国際機関と協力しています (図 11)。SSRN もその協力機

**Elsevier and open access**

Gold open access	Green open access
<ul style="list-style-type: none"> <li>• 2nd largest gold open access publisher</li> <li>• Publish over 170 fully open access journals</li> <li>• Publish over 1,850 hybrid journals</li> <li>• Published over 25,000 open access articles in 2016</li> <li>• Choice of either a commercial (CC BY) or non-commercial (CC-BY-NC-ND) user license.</li> <li>• Article publishing charges (APCs) range from \$150- \$5000 (US Dollars)</li> </ul>	<ul style="list-style-type: none"> <li>• Largest publisher enabler of green open access</li> <li>• All 2500+ journals provide a green open access option</li> <li>• Participate in CHORUS and support pilots with institutions and international funders</li> <li>• Free API program to fuel repositories</li> <li>• Share link service provides 50 days free access to recently published research</li> <li>• Open archives in 108 journals, including Cell Press titles after 12 months</li> </ul>

(図 7)

**Storing research data**

- Submitting authors can store research data in Mendeley data repository and link this to their article on ScienceDirect
- Open access, free-to-use repository
- Hosts all formats (raw, processed data, tables, code and software)
- 10GB free storage and CC0 license option
- Assigns data DOI to make it independently citable
- Partnership with DANS for long term storage and preservation

(図 9)

**Fueling Institutional repositories: Free API program**

Get metadata & abstracts	Show the final version	Keep users on your IR	Make manuscripts public
Search API can extract the metadata and abstracts of your affiliated authors to help populate your repository, enhance discoverability and create links or embed final version	Entitlements API** can ensure subscribed users see the final version and visitors can see the accepted manuscript hosted on your repository	Article Retrieval API embeds the final article in your repository so subscribers don't leave your site to access the article	Hosting permissions API can retrieve embargo end dates so you can make manuscripts automatically available to the public after embargo

**Scopus** Did you know? Scopus customers can also integrate Scopus APIs to retrieve metadata and abstracts across all publishers

Please Note: \*\*Minimum part of the program. //dev.elsevier.com/tecdoc\_sd\_ir\_integration.html

(図 8)

**Open peer review reports**

Pilot in 5 journals to experiment with publishing peer review reports as openly available research outputs with their own DOI.

Editors	Reviewers	Authors
<ul style="list-style-type: none"> <li>• 70% of those surveyed believe that reports are more in depth and constructive i.e. improved and more useful to authors, as a result of being made open</li> <li>• Some concern it may make it harder to find reviewers.</li> </ul>	<ul style="list-style-type: none"> <li>• Like the ability to get recognition</li> <li>• Over 95% said publishing review reports didn't influence their recommendation nor would it influence them accepting further review invites for the journal.</li> </ul>	<p>Generally positive about having open peer review reports next to their article</p> <p>Little impact on their journal choice, indeed slightly more likely to choose a journal offering this.</p>

Forward look: Extend to other interested journals and provide more data about the peer review process on ScienceDirect

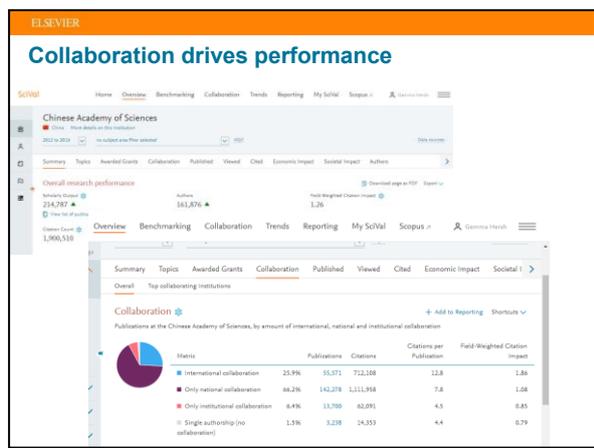
(図 10)

関の一つになっていて、SSRN の利用者、論文、著者の数からそのことが分かります (図 12)。Mendeley はもう一つの協力例で、参考文献やその他の文献を初期段階で共有するものです。他者との協力という点では、社会的影響を評価する Atlas Award もあります。AudioSlides は、他分野の専門家や研究者に動画や 5 分間のプレゼンテーションを提供します。図 13 は、エルゼビアの最新の報告書「Gender in the Global Research Landscape (世界の研究環境におけるジェンダー)」です。SSRN が今週、初期段階の研究共有に向けて Women's & Gender Studies Research Network を立ち上げたことが示されています。

**より透明に**

次に透明性ですが、私たちも他の多くの出版社にならって、科学コミュニティとともに研究を透明性のあ

るオープンなものにするためのデータガイドラインを採用しています。また、私たちはデータベースの Scopus で、検索をよりしやすくしました。とても重要なツールである Scopus が、もっと幅広く使えるように改善されたのです。DataSearch は面白いツールで、メタデータを検索するだけでなく、自分でデータそのものを検索できるので、全体をダウンロードする前にそのデータセットに関心を持てるかどうか、判断することができます。また、私たちはデータの引用を増やすために、FORCE 11 とも協働しています。データの引用は、論文の引用とは種類が異なりますが、私たちはデータセットの作成者が認識されることも重要だと考えています。そして、ネガティブな結果となった情報の報告も強化しました。そうすれば他の人たちの研究から学べるからです。さらに、データだけを掲載する専門ジャーナルもあります。最初にお伝えしたとおり、Elsevier はデータインフォマティクス企業になりつつあります。



(図 11)

**SSRN – tomorrow’s research today**

- SSRN is devoted to the rapid worldwide dissemination of research and is composed of a number of specialized research networks.
- SSRN’s eLibrary provides 761,906 research papers from 354,694 researchers across 30 disciplines.

**Tomorrow’s Research Today**

SSRN’s global reach provides the tools to disseminate innovative ideas through our online repository, making librarians’ jobs easier and keeping academics up-to-date.

**SSRN provides the tools; your academic team enjoys the benefit.**

**Always Growing. Always Something New.**

- 2.2M+ Users
- 800K+ Papers
- 326K+ Authors
- 107M+ Downloads
- 9.1M+ References
- 5.8M+ Citations
- 8.9M+ Footnotes

SSRN  
Tomorrow’s Research Today

(図 12)

**A more inclusive world of research**

Gender in the Global Research Landscape Report

“As a steward of world research, Elsevier has a responsibility to promote gender equality in science, technology, engineering, and mathematics and advance understanding of the impact of gender, sex, and diversity in research. In this regard, Elsevier fully supports the United Nations’ Sustainable Development Goal 5: “to achieve gender equality and empower all women and girls,” and the Global Research Council’s Statement of Principles and Actions Promoting the Equality and Status of Women in Research.” – Ros Mohed, Elsevier CEO

Critical issues related to gender disparity and bias must be examined by sound studies. Drawing upon our high-quality global data sources.

(図 13)

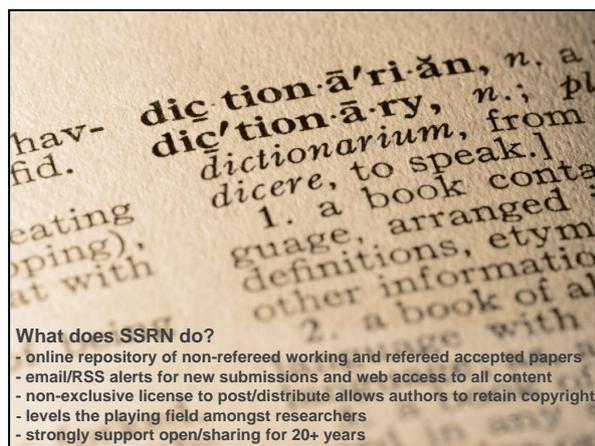
**SSRN について**

SSRN は 22 年ほど前にスタートしました。図 14 がその基本情報です。これらのポイントのうち一番重要なのは四つ目の、研究者に公平な競争の場を提供するという点だと思います。ハーバード大学でも、私が住むニューヨーク州ロチェスターのモンロー・コミュニティ・カレッジでも関係ないということです。出身の国や地域、どの言語で研究を行うかも関係ありません。

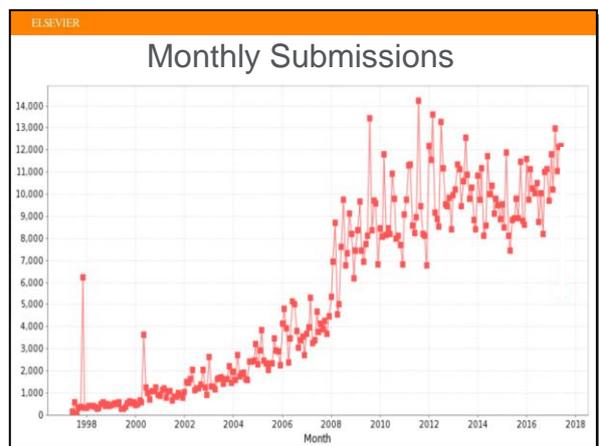
SSRN には約 35 万人の著者がいます。図 15 のグラフで急に飛び出ている部分は、他の組織が SSRN に参加したときです。図 16 のグラフは論文投稿の数です。同様に飛び出ている部分があります。また、ここ数年で大幅に増加しました。図 17 は重要なスライドだと思います。青の点は 1 日の投稿数で、赤の点は 1 日の修正数です。2007 年 11 月に、私たちは SSRN ライブ

ラリの論文の修正やアップデートを認めるようになりました。このグラフは 2012 年 4 月で終わっています。2012 年 4 月のこの日には、新規投稿より修正の方がたくさんありました。これは、SSRN の発展において重要な日です。SSRN が生きている、呼吸している情報の集まりであることを意味しているからです。私たちは、時間とともに研究が発展し、より良くなっていくことを望んでおり、早期の共有によってその速度は早くなります。

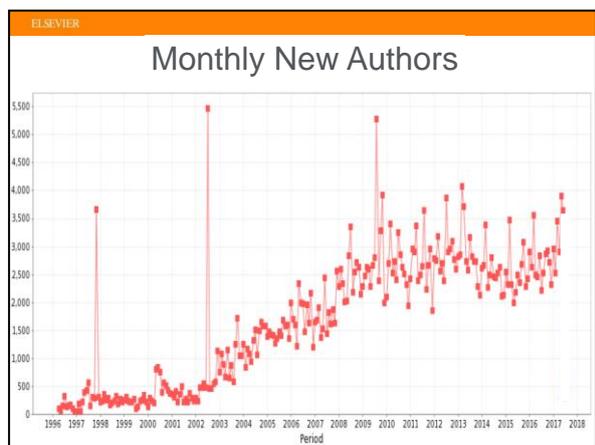
初期段階の研究にアクセスできれば、利用できる、自身に関係のある研究が増えます。研究者のアイデアや草稿、ワーキングペーパー、会議録、プレプリントを、最終的に発表された論文と同じように利用することができます。これまでは一つのバージョンしかなかったものが、いまや三つのバージョンがあります(図 18)。さらに、発表後のバージョンも追加し、研究を



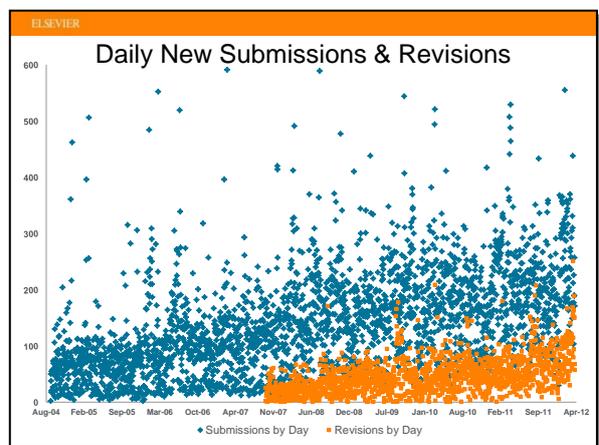
(図 14)



(図 16)



(図 15)

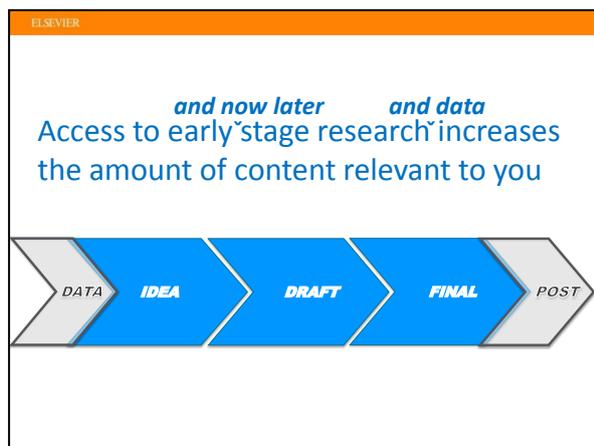


(図 17)

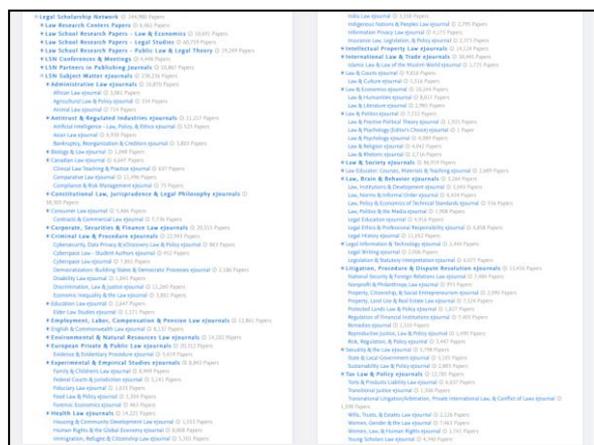
アップデートできるようにしました。データも追加すると、一つの研究論文から最高で五つのバージョンが出せるようになりました (図 19)。これらのバージョンの管理は複雑です。図 20 は SSRN 内の法学研究分野でのその複雑さの一例を示したものです。



(図 18)



(図 19)



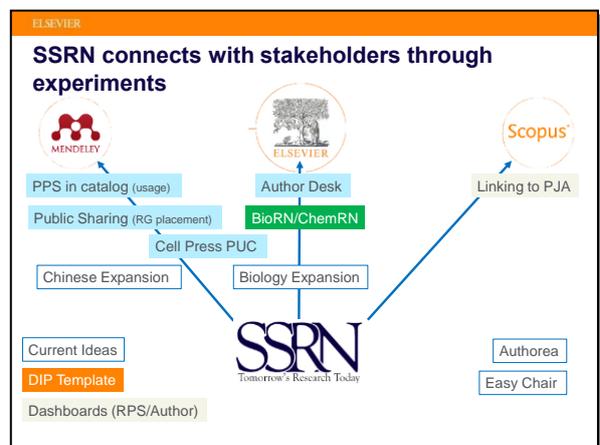
(図 20)

SSRN は、社会科学でスタートして、人文科学、生物学、化学が追加され、今週には女性・ジェンダー研究が追加されました。来月には工学を追加します。私たちのスローガン、マントラは、「Tomorrow's Research Today (今日、明日の研究を)」です。2014 年の「Journal of Financial Economics」に発表された論文は、一つを除いて全て、3 年半前には SSRN の eLibrary データベースで発表されていました。世の中に目を向けると、パートナーなしに私たちの運命を全うすることはできないということが分かったので、Elsevier と提携することにしました。それによってギャップを埋めることができたと考えています (図 21)。今、私たちは「Tomorrow's Research Today」から出発して、研究者をつなぎ合わせ、アイデアを展開させ、初期の研究をシェアする手助けをしているのです。

私が考えるイノベーションのシンプルな定義は、よ



(図 21)



(図 22)

り幅広い、より深いものに触れることで、新しいものを生み出すということです。私たちは、SSRN が最も幅広く、最も深い一連の研究を提示できるようにしたいと考えています。これは、研究パフォーマンスを改善したり、研究者が素晴らしい、新しく革新的な研究をもっと素早くできるよう支援したりするためです。私たちはギャップを埋めてきた一方で、実験やパイロット試験で実際にいろいろなピースをつなぎ合わせて、研究をもっと先に進めようとしているのです (図 22)。私は最近、SSRNblog.com のブログにオープンアクセスウィークを祝う投稿をして、過去 15 年間でオープンアクセスがどのように発展してきたか論じました。オープンアクセスの良い面、オープンアクセスのためにもっとうまくやらなければならないことについても論じています。ぜひご覧ください。

**中国、韓国、日本での SSRN 利用統計**

私はこの 1 週間半で北京、ソウルを訪ね、東京にやってきました。私たちは、いろいろな学校でさまざまな研究者が SSRN を活用する方法を見てきました。この数字は、中国 (図 23)、韓国 (図 24)、日本 (図 25) での利用に関する統計を示したものです。研究者に、母国語で執筆し、研究し、共有できることを知らせるために、もっとやらなければならないことがあるということが分かりました。また、私たちは研究者に、SSRN は機関リポジトリ、分野ごとのリポジトリ、国別リポジトリとしても使えるということを知らせる必要もあります。

●フロア 1 バイオサイエンスデータベースセンターの職員です。中国と韓国と日本のユーザー数の比較で、日本のユーザー数がとても少ないように見られました。その理由は何だと分析されていますか。

●Gordon 他の講演者の方々とのランチのときにその

話題が出ました。中国の方がコンテンツは多いですが、SSRN を使いはじめたのは日本の方が先です。日本の研究は他の 2 カ国に比べて、早くからもっと洗練されていたと思います。私は理解を深め学ぶためにここに来ているわけですが、私たちは世代交代の時期に差し掛かって

Source: SSRN Website

School / Department	Authors	Papers	Downloads
The University of Hong Kong - Faculty of Business and Economics :: China	96	750	201,625
Tsinghua University - School of Economics & Management :: China	105	470	108,182
Peking University - Guanghua School of Management :: China	89	360	92,706
Cheung Kong Graduate School of Business :: China	21	163	84,501
China Europe International Business School (CEIBS) :: China	46	268	80,546
University of International Business and Economics (UIBE) :: China	102	245	26,800
Fudan University - School of Management :: China	50	137	22,527
Peking University - Shenzhen Graduate School of Business :: China	27	117	20,118
Shanghai Jiao Tong University (SJTU) - Antai College of Economics and Management :: China	49	123	18,730
Peking University - HSBK School of Business :: China	13	65	18,535

(図 23)

Source: SSRN Website

School / Department	Authors	Papers	Downloads
Korea University Business School (KUBS) :: Korea	53	221	76,540
Seoul National University - College of Business Administration :: Korea, Republic Of (south Korea)	30	150	57,650
KDI School of Public Policy and Management :: Korea, Republic Of (south Korea)	57	201	30,297
KAIST Business School :: Korea, Republic Of (south Korea)	48	128	22,249
Yonsei University, Seoul Campus, College of Business and Economics :: Korea	21	122	14,444
Korea Advanced Institute of Science and Technology (KAIST) - College of Business :: Korea, Republic Of (south Korea)	47	112	11,653
Sungkyunkwan University - School of Business Administration :: Korea, Republic Of (south Korea)	12	35	8,722
SolBridge International School of Business :: Korea, Republic Of (south Korea)	9	23	5,742
Gachon University - College of Business and Economics :: Korea	17	64	4,749
Hankuk University of Foreign Studies - College of Economics and Business :: Korea	8	19	4,018

(図 24)

Source: SSRN Website

School / Department	Authors	Papers	Downloads
Tokyo University of Science - School of Management :: Japan	4	30	7,911
Kobe University - Research Institute for Economics & Business Administration :: Japan	7	43	5,921
Kobe University - Graduate School of Business Administration :: Japan	21	42	4,894
Hitotsubashi University - Graduate School of Commerce and Management :: Japan	16	35	3,490
Hokkaido University - Graduate School of Economics & Business Administration :: Japan	17	57	3,331
Aoyama Gakuin University - Graduate School of International Management :: Japan	3	13	2,531
Hosei University - Graduate School of Business Administration :: Japan	1	13	2,432
University of Tsukuba - Graduate School of Business Sciences :: Japan	3	8	2,371
Keio University - Faculty of Business and Commerce :: Japan	10	30	2,143
Keio University - Graduate School of Business Administration :: Japan	4	13	1,986

(図 25)

いるのではないかと思います。以前はより先進的な日本人研究者の多くが SSRN を使っていました。今は、もっと若く、頭角を現している研究者が SSRN を使うようになってきています。私は、彼らができるだけ早く成功を収められるよう支援するためにやってきました。

●フロア 2 横浜国立大学の深貝と申します。私は SPARC Japan の運営委員の一人です。商業ベースの仕組みと、arXiv.org のように学者・研究者が始めた仕組みはなじまないところがあるかもしれないということ、これが今後どうなるのかは大切だと思います。

今お話しになったように、世代交代があれば、SSRN、あるいはパブリッシャーが提供するものを使う人の比率が増えるのは確かだと思います。それは便利なものになじむ世代が増えるということですが、それで学術がうまくいくかという別の問題です。産業革命から現代に至るまで、いろいろな技術革新がありました。単に新しいものを使うことが大事なのではないのです。大事なのは、そこからどういう新しい文化をつくれるかということです。今、生活において、世の中の在り方において、新しい技術の意味が問われています。学術においてもどうかということをお互に考えなくてはなりません。

オープンサイエンスの中でのツールは誰が提供しているのか、自分たちでつくっているのか。商業ベースで提供していただけるのは確かに便利であって、ありがたいですが、その中でどういう研究が育ち、その知識がこの世の中においてどういう貢献ができるかということは、それはそれで別に考えなければいけないのではないのでしょうか。

●Gordon とてもいい質問ですね。少し話を絞ってお答えしましょう。私は、単なる商業対学術という見方はしていません。私たちは、少なくとも各国の政府と資金援助団体を巻き込んでいかなければなりません。

私の友人に Simeon Warner と Paul Ginsparg がいます。Simeon は、arXiv が持続可能なビジネスモデルを模索

していたとき、ロチェスターにやってきました。何に、どう課金できるか、そして一番重要なこととして、どう自ら資金を調達するかについて、私たちはたくさんのシナリオを検討しました。商業面の問題は、bioRxiv や Open Science Framework などを見てみるともっと複雑です。bioRxiv は現在、フェイスブックを創業した Zuckerberg の Chan Zuckerberg Initiative (CZI) の資金提供を受けています。Open Science Framework は、億万長者のヘッジファンドマネージャーが運営する Laura and John Arnold Foundation の資金提供を受けています。

私たちの選択肢の一つは、政府です。政府は有権者、党员、特別利益団体に対してその活動の正当性を説明しなければなりません。他の選択肢としては、bioRxiv や Open Science Framework のように、億万長者のビジネスマンが非営利の資金提供を行い、このような手段を通じて意思決定をしてくれるということがあります。三つ目の選択肢は、営利組織です。彼らは価値を提供できなければ、顧客を失ってしまいます。私は、このうちのどれが本質的に良いとも悪いとも思いません。どれも欠点があります。それぞれに欠点があるということを認めない人にはいらさらされてしまいます。

私は、SSRN が Elsevier に加わったことにワクワクしています。Elsevier は巨大な営利企業で、良いことばかりではありませんでした。この共有というゲームは非常に高くつくものになっていて、自力では成功できない、と私たちは分かっていました。幸運なことに、自分たちがしていることの可能性を信じ、全面的にサポートしてくれる大きな組織と提携することができました。ふさわしいパートナーがいなければ、生物学や化学、女性・ジェンダー研究、工学の全てを今年追加することはできなかったでしょう。私としては、心変わりして、資金提供をやめてしまう組織に頼るよりも、みんなが去ってしまうような状況でも価値を提供し続けなければならない組織を信頼しています。私たちはただ、SSRN はしっかりと管理されており、情報のセ

キュリティが確保されているということを保証しなければなりません。それが、ずっと長期にわたって仕事を続けられるようにするために、私が SSRN の中で最善を尽くそうと努力してきたことです。でも、私はこの週末東京に来ていいお天気を楽しんでいますけれど、それと同じでどこにも保証はないのです。

●フロア3 NISTEP の林 (SPARC Japan 運営委員) です。2013 年に、Fiesole Retreat という出版社を中心とした非公式会合があり、セッションは違いましたが、Gregg さんと一緒に講演してディスカッションする機会がありました。そのときのエピソードを交えてシンプルな質問があります。

2011 年に SPARC Japan セミナーで Mendeley の CEO を呼んだ際に、「この先、どうするの。もしかして Elsevier が買ってしまったらどうしよう？」と話したところ、彼は明らかに反応が変わり、その後、Mendeley は Elsevier に買収されました。

2013 年に私は Gregg さんに「ビジネスサステナビリティはどうされるのですか。商業的な資本が入る可能性はありますか？」と質問しました。そのときは、SSRN としてブランドを大事にしていきたいというお話をされましたが、今は Elsevier の傘下にあります。

私は非常にニュートラルに、今まさにおっしゃっていたように誰も将来が見通せない中、いろいろチャレンジをしていく上で、パトロンが誰であるかが非常に大事であるということを理解しています。

その上で質問が二つあります。一つ目は、SSRN が Elsevier の中でどうサステナブルかということです。二つ目は、今の立場から今の arXiv.org のビジネスサステナビリティはどのように見えるかということです。Taxation model と呼ばれることもあります。会員制で、会員からお金を集めて、その代わり公共にフリーに出すというモデルについて、ご自身の見通し、パースペクティブを教えてください。

●Gordon まず、Mendeley と SSRN にはいろいろな

選択肢があったことを言うべきだと思います。正直なところ、資金的にはもっといいオファーがあったのですが、SSRN の発展にとっていいものではありませんでした。この部屋の中で、20 年以上同じ仕事をしている人は何人いますか？ほんの少しだけですね。それだけ長い間何かができるということは、大好きだからやっているということです。いろいろな問題、いろいろなフラストレーションにもかかわらず、続けていきたいわけです。私たちは、できる限り SSRN を今のまま続けていくことを保証できる取引を構築しました。

客観的に見てみると、Elsevier もその他の全ての出版社も変化していることが分かります。愚かな人たちではありません。秘密の部屋で意思決定をしているわけではありません。あらゆることがオープンなところで起こっています。みんながいろいろなことをしていて、私たちの誰も、20 年後、10 年後、あるいはたった 5 年後でさえ、未来がどうなっているか分かりません。私は自分のチームや、Elsevier 内の他の人たちにも、「今はクレイジーな時代だ。それを受け入れるか、そうでないなら違うことをしなさい。でも、重要なのは間違った答えは存在しないということだ。このことを毎日理解しておく必要がある」と言っています。

最初のご質問に戻りますと、私は学会やその他の会員制組織と過ごす時間が多いのですが、皆さん会員の維持に苦労しています。会計学を学んだ身としては、このようなビジネスモデルは採用しません。ですが、毎年何十万マイルも飛び回って、あらゆる場所でここにいる皆さんのように賢い人たちの話を聞いている身ですから、私が間違っている可能性も大いにあります。もっと大切なことは、このような方々の成功を、私は心から願っているということです。

## 第2回 SPARC Japan セミナー2017

「プレプリントとオープンアクセス」

### 化学分野におけるプレプリントの位置付け・課題等について

生長 幸之助

(東京大学大学院薬学系研究科/化学ポータルサイト Chem-Station 副代表)

#### 講演要旨



現行の査読システムに諸々の問題点が見えつつある状況下において、迅速な研究情報共有プラットフォームとしてのプレプリント投稿が各分野で台頭しつつある。化学分野でも、アメリカ化学会主導でChemRxivというプレプリントサーバの運営が先日より開始され、注目を集めている。本講演では、プレプリント投稿が今後化学分野でどのような普及・活用・発展を遂げていくのかについて、事例を参照しつつ議論してみる。演者は大学にて教員・研究者として働く傍ら、化学分野の国内最大二次情報プラットフォーム Chem-Station の運営にも長年関わっている。研究者とメディア運営の二足のわらじを履く立場からの視点も併せて述べてみたい。



#### 生長 幸之助

1980年徳島生まれ。2003年 東京大学薬学部 卒業。2008年 東京大学大学院薬学系研究科 博士課程修了(指導教員:柴崎 正勝)。同年 博士(薬学)。2008-2010年 カリフォルニア大学ロサンゼルス校 化学/生物化学科 (Omar M. Yaghi 研究室) にJSPS海外特別研究員として赴任。2010-2016年 東京大学大学院薬学系研究科 助教。2016年より現職(講師)。化学ポータルサイト Chem-Station 副代表(2002-)、ERATO金井分子触媒生命プロジェクト研究総括補佐(2015-2017)を兼任。

私の名前は生長(おいさき)と読みます。大学で教員を務めていますが、本業は実験化学の研究者です。バイアルやフラスコに化学物質を混ぜてものをつくるということをやってきました。自分なりの立場からこのプレプリントについての見方や意見、将来どうなっていくかということをお話できればと思います。

#### 1.自己紹介

私の専門は薬学の中でも有機合成化学という分野です。もっと狭い話をする、触媒反応開発という分野になります。一番皆さんになじみのある研究者の名前を挙げると、2010年にノーベル賞を取られた、クロスカップリングの鈴木先生や根岸先生の方針が一番近いです。この薬という複雑な構造を持った有機化合物に対して、もっと良いものにしたい、どうやったら効

く分子になるかということを考えていろいろ加工していきます。

これを一度にやろうとすると大変なのですが、魔法の粉といいですか、特殊な化学反応を進行させるような触媒を自分たちの手でテーラーメイドし、ショートカット反応を起こして薬を効率的に供給したり、今までアクセスできなかった構造の化合物にアクセスできるようにしたりしようという発想で研究をしています。この触媒を開発するというのが私の主な研究テーマの一つです。長い目で見ると、薬開発を促進させるというビジョンを持ってやっています。

図1は、触媒開発の一環で出てきたごく最近の成果で、一番分かりやすいものです。今は小さな有機化合物だけではなく、抗体なども薬になっています。かなり大きなサイズの分子も薬になっています。タンパク

質なども薬になるという時代になっているのですが、これを人間の力で加工することができれば非常に面白いのではないかと、そういう化学反応を自分たちの手で開発しようというテーマにも取り組んでいます。

最近、触媒の基になる試薬がタンパク質のアミノ酸であるトリプトファン残基にくっついて、トリプトファン選択的に化学反応を起こすというものを開発しました。これを使って抗体薬を化学修飾して、もっと強い効き目を持つ抗体薬に変えるという研究にも現在取り組んでいます。

このように化学反応ベースで新しい薬に道を付けようという考え方で研究をしている人間です。

## 2.Chem-Station について

### 2-1.運営体制

ここまでが研究の話なのですが、これ以上に私は長年 Chem-Station という化学のサイトの運営に携わっています。

Chem-Station は図 2 のような見た目のウェブサイトです。「ケムステ」と検索を打てば出てきます。これは一体どういう化学サイトかというと、日本最大のアクセス数を稼いでいる、化学の情報を集約してまとめて発信する、いわばポータルサイトと呼ばれるものです。

設立は 2000 年で、私は 2 年目ぐらいから、約 16 年研究室に入る前からずっと運営に関わっています。このサイトは「化学とウェブの融合」という大きな目的

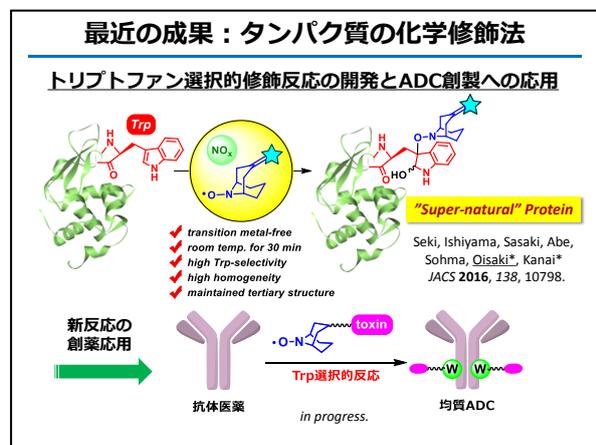
を掲げていて、将来的に化学とウェブを融合させることができないかということを考えて日々やっております。具体的には、ウェブにたくさん存在している化学に関する情報を分かりやすい形でまとめて提供することをはば人力でやっている形になります。

これだけ長く続けているので、知名度も最近すごく上がってきて、アクセス数が 80 万~240 万 pv/月となっています。

公開ページ数も、これは蓄積の産物だと思うのですが、5,000 ページを超えていて、ボランティアも含めていろいろなスタッフの方にご協力いただいています。パーマネントのスタッフはおらず、本業がある中で協力いただいています。

運営も、ある程度持続可能なものにしたいため、アクセス数があるページなので、スポンサーから広告を出してもらって、サーバー代、記事のお金に換えていくというようなやり方をしています。化学系の会社と長いお付き合いがあります。

私は副代表という位置付けで、代表は早稲田大学理工学術院准教授の山口潤一郎さんです。彼がサイトを立ち上げて、私が後で加入しました。二人が一番古株なので、代表、副代表という関係でずっとやっています。最初はスタッフの数は 2 人ぐらいしかいなかったのですが、年を追うにつれて知名度が上がって、やりたいという人も募れて、今ではスタッフの数が 100 名を超えています。他にもサポートしてくださっている方がたくさんいます。



(図 1)



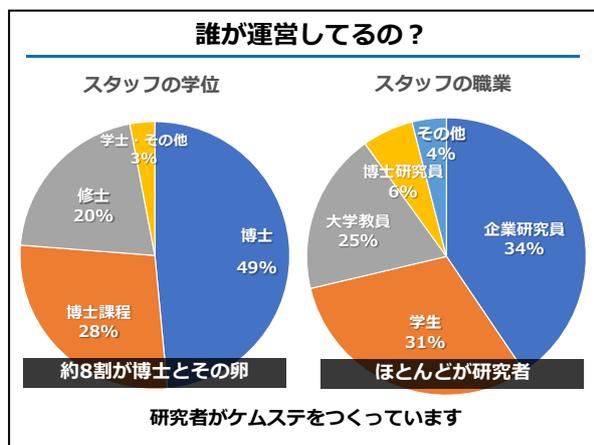
(図 2)

このサイトを運営しているのはどういう方かというと、基本的には皆さん本業があって、ボランティアに近い形でやってくださっている方が多いです（図 3）。これはかなり特徴的かと思うのですが、実際にスタッフがどういう学歴を持った方かということをもとめ直してみると、博士課程の学生か、博士号を既に持った方が約7割を占めて、そういう人たちが化学のウェブ上の記事や専門的な文献などを読んでまとめて日本語で記事をつかってオープンにするという作業に関わっています。スタッフの目がしっかりしているということで、信頼性も担保されています。

もともと研究的な視野から出てきているサイトなので、研究者に役立つ情報を分かりやすく提供しようという考え方がかなりあり、必然的にスタッフの本業も研究であることが多いです。端的に、ケムステをつくっているのは研究者であると言うことができます。

## 2-2.発信内容

研究者がどういう情報を発信しているのかというと、実は研究者専門というわけではなく、一般市民にも分かるぐらいのところまでカバーして、スタッフの書きたいようにやってもらっています（図 4）。研究者ならではの視点からその情報を読み解いて、分かりやすく文章にして発信するブログを書いたり、ニュースの形で紹介したりします。ニュース自体は一般の新聞が取り上げているものなのですが、それに例えば化学者ならではのコメントを付けて、解釈して出してあげる



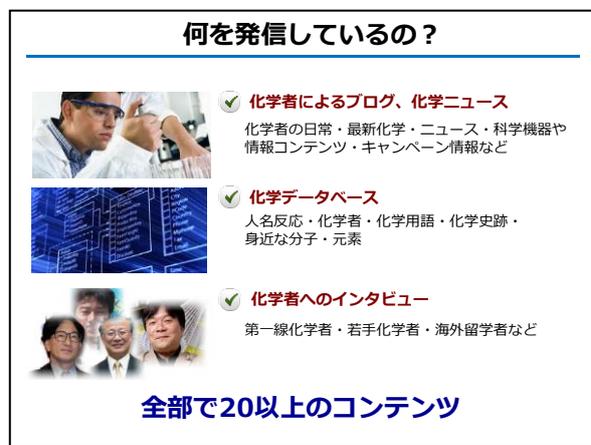
(図 3)

ことで付加価値が出ています。

こちらも特徴的ですが、化学者ならではの必要な情報を整理してまとめ上げたデータベースをつくっています。かなりアクセス数があります。また、化学者がつくっているサイトということで、有名な化学者はどうしているのかを考えて研究に取り組んでいるのか、毎日どうやって過ごしているのかという生の声をお届けするようなコンテンツもあります。サイエンスコミュニケーションの走りのようなことをずっとやってきて、今そのような形で認知されているというサイトです。今は大変バラエティに富んだコンテンツが提供されています。

ある化学情報を、日本人に分かりやすいように日本語で発信するというところに一番ニーズがあります。しかし、外国にも届けないと、日本の化学がいくら良いものでも、全体として知名度が上がらないですし、外国からクレームが付いたときに反論できる土壌がない、そういうことが実際に別の件であったので、そこに問題意識を向けました。基本的には日本語のサイトを英語化して、中国人スタッフを抱えて中国語に翻訳して、二つのサイトを並走させながら日本の化学というものを世界に発信していくということもしています（図 5）。

また、サイエンスコミュニケーションの一環だと思いますが、化学が一番注目を集める時期はノーベル賞受賞の時期かと思います。それに合わせて、ChemStation が旗を振って、誰がノーベル化学賞を取るのか



(図 4)

という予想投票企画を Facebook の機能を使って行っています。今年は 500 票ぐらいアンケートが集まりました。Facebook を使っている人が気軽にワンクリックで誰がノーベル化学賞を取るかいうことを投票できるような企画です。当たったら賞金が出ます。

それと並行して、ノーベル賞が発表されたら、どういふ人たちがノーベル賞をどういふ業績で取っているのかということ日本語の記事にしています。

普通の日本のメディアだと、日本人が取らない限りは情報がなかなか出てこないのです。私たちはそれはどうかと思っていて、ノーベル賞を取った素晴らしい業績なのだから、ぜひみんなにも知ってほしいということで、今年はクライオ電子顕微鏡について、自分の勉強も兼ねてきちんと調べて、日本の記事として書いています。それが巡り巡って化学を皆さんに知ってもらうような活動になっているのではないかと信じて、毎年やっています。

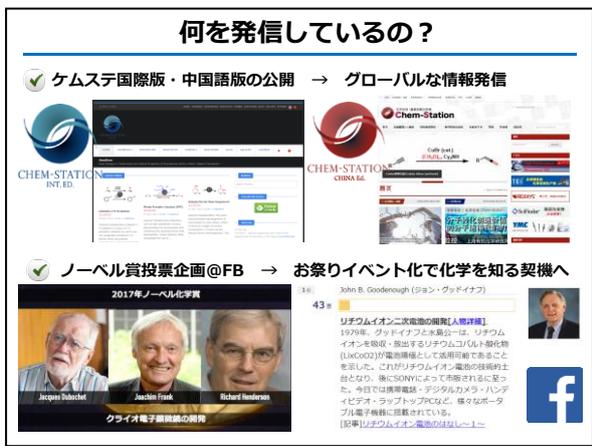
基本的にはそのようにウェブベースでの活動をしているのですが、リアルの活動にも少しずつ進出しはじめています。図 6 の上は、科学技術振興機構が開催している、「サイエンスアゴラ」というサイエンスを土台にしたお子様向けのお祭りです。それに Chem-Station のブランドで参加し、化学実験のブースをつかって、化学は面白いということをお子さんやそのご両親に体験していただくということをスタッフ主導でやっています。

また、このケムステというサイトは若い人がかなり

見てくれていることが特徴です。30 代以下が読者として多く、それに日本化学会が目をつけました。日本化学会年会は毎年行われている学会で、懇親会も開いています。若者向けがない、開催してもおじいさん、おばあさんしか来ないという実情があったようで、何とか若者を引き留めたいということ、当時会長だった玉尾皓平先生が考えたのだと思います。そこでうちの代表に話を持ってきて、ぜひケムステ主導で若者の懇親会をつくってほしいということで実際に実現したのが「ケムステイブニングミキサー」というものです。下の写真は今年 3 月の懇親会の様子で、このときは 500 人ぐらい集まりました。

なぜこんなに集まったかという、マグロの解体ショーを代表が自腹を切つて企画したからということがあります。これを見たいがためにいろいろな学生たちが来て、大盛況になっていました。これは代表の企画力がすごかったという一例かと思ひます。

というわけで、Chem-Station のようなサイトにはほとんどありません。本当に唯一無二と言ってもいいぐらいのレベルでやっているとわれわれは自負しています。その中でグローバルな発信も視野に入れながら、みんな本業がある中でやってくれているので、そこまで進むかどうかという問題はありますが、できれば世界一の化学メディアを目指していきたいです。社会に化学の情報を行き渡らせて、化学の好きな人を増やしていきたいというビジョンを持って続けております。



(図 5)



(図 6)

### 3.化学とプレプリントサーバー

#### 3-1.化学系プレプリントサーバーChemRxiv

ここから今日の本題に入ります。ケムステという活動をしていて、プレプリントサーバーというものに個人的に興味を持ち、まとめて紹介記事を書きました。これがプレプリントサーバーについてここで話すことになったきっかけです。

最初に書いたのは、化学系プレプリントサーバー ChemRxiv の設立が決まりましたというイントロ記事です (図 7)。これはまだできていない段階で取り上げて、化学のみんなはプレプリントサーバーを多分知らないだろうと思ったので、プレプリントサーバーとはこういうものなのですよという追加の記事を付けました。

実際にベータ版ができるということが分かり、やりはじめましたということニュース記事にして、ばら

ばらと見てみた感じの注意点などもまとめたブログ記事を研究者向けに出しました。この ChemRxiv が化学系のプレプリントサーバーとして米国化学会の手によって設立されて運用が始まったのが 2017 年 8 月の話です (図 8)。

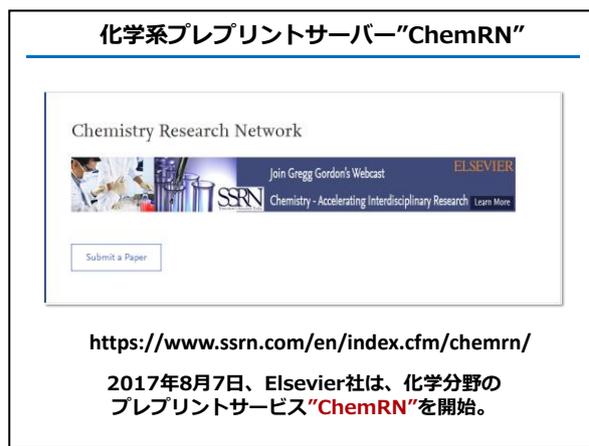
先ほどの Gordon さんのお話に出てきたような、SSRN 主導で化学分野のプレプリントを集めるということも今まさに始まらんとしています (図 9)。こちらでも 2017 年 8 月にアナウンスが出たところです。

SSRN のことは横に置いておいて、ChemRxiv に話を絞っていききたいと思います。やはり始まったばかりの取り組みで、活発に ChemRxiv がコミュニティで使われているのかと言われると、まだまだこれからという実感はあります。

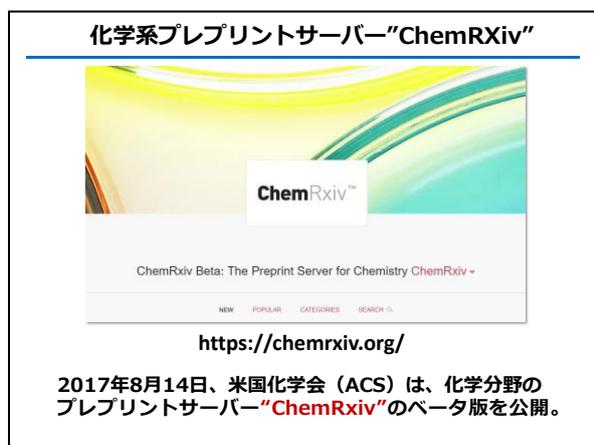
図 10 は ChemRxiv のサイトから引用したスタッツです。ChemRxiv の 2017 年 8 月 10 日から 10 月 15 日ま



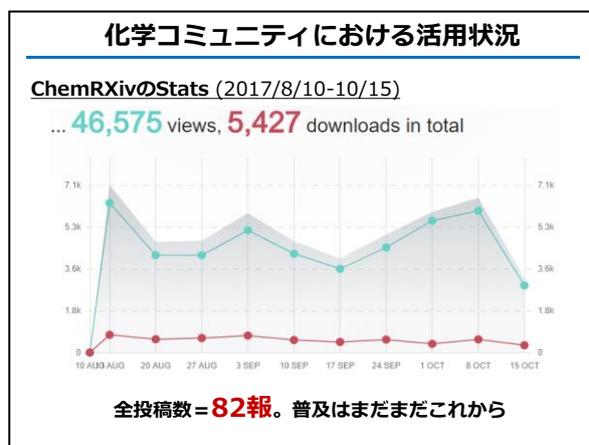
(図 7)



(図 9)



(図 8)



(図 10)

でのアクセス数は 46,575、ダウンロード数は 5,472 です。この短い2カ月の期間で、私が確認した最近の論文投稿数は全部で 82 報で、arXiv.org の 11 万報とは雲泥の差です。これはできたばかりなので仕方がない、まだまだこれからだご理解いただくしかないと思います。

### 3-2.プレプリント投稿のメリット

プレプリント投稿の大きなメリット（図 11）は、とにかく査読過程がスキップできて、素早い情報交換が行えるということです。ただ、査読を終えていないので考えて使ってくださいということです。また、誰が最初にその仕事をしたかということが査読を経ている間にうやむやになっていかなないということが大きなポイントかと思えます。査読中に、アイデアの盗用ということが問題になると思うのですが、ある程度はそれも防ぐことができるということです。そして、オープンアクセスのモデルなので、ユーザーにとってはコスト的にありがたい取り組みです。誰でも最先端に近い情報が広く手に入るということで、よいところがあるため、化学分野でもこれをやっていこうという流れになりつつあります。

査読期間については、今まで化学の分野ではそれほど問題になっていませんでした。追試がしやすく、バイオ系に比べたら実験期間もそれほど長いわけではないので、リバイズもそこまで時間がかからなかったのです。けれども、最近様変わりしてきているのではな

### プレプリント投稿のメリット

- ✓ **素早い情報交換が行える**  
 時間のかかる査読過程をスキップ  
 (※未完成品であることには留意)
- ✓ **先取権の明確化**  
 revisionもタイムスタンプで管理  
 査読者によるアイデア盗用の抑止力にも
- ✓ **編集・出版・購読コストの抑制**  
 オープンアクセスモデル  
 開かれた知の共有へ

(図 11)

いかという実例を挙げたいと思います。図 12 は私が関わった論文で、ファーストオーサーが ERATO のファンドで雇われていたポスドクの方で、星が付いている方はグループリーダーです。ERATO は 5 年間のファンドなので、5 年後には 2 人ともどこかに職を探さなければいけません。当然ながら追い詰められながら仕事をします。3 年ぐらいのときに論文が出せそうだということで、Nature Chemistry に投稿しました。そこからレフェリーに相当シビアなリバイズ要求を送られました。リバイズ 3 ぐらいまで行ったと思うのですが、結果として、レシーブが 2014 年 9 月、アクセプトが 2016 年 5 月で、ほとんど 2 年近い期間がかかってしまい、最終的に論文を通ったのがほとんど任期ぎりぎりという、考えたくもないような状況になりました。

なぜこうなってしまったかということ、この仕事は純粋な化学だけではなく、細胞を使うなど、バイオロジーと融合した面もあり、時間のかかる実験がたくさんあって、リバイズも時間がかかってしまったからです。要求される実験量も多いという事情が化学の世界にもたくさん出てきたことが、査読の時間が長くなっている背景にあるのではないかと個人的には考えています。

図 13 は、合成化学のトップラボのスクリプス研究所の研究者が投稿した、複雑な化合物を斬新な考え方で化学合成したという論文です。

図 14 は論文を ChemRxiv で見たときのブラウザ画面です。論文テキストがあって、下の方にダウンロード

### 化学分野でも査読期間の長期化が懸念

【演者の関わった実例】

ARTICLES  
PUBLISHED ONLINE: 27 JUNE 2016 | DOI: 10.1038/NCHEM.2550

**nature  
chemistry**

**Switchable photooxygenation catalysts that sense higher-order amyloid structures**

Atsuhiko Taniguchi<sup>1,2</sup>, Yusuke Shimizu<sup>1</sup>, Kounosuke Oisaki<sup>1</sup>, Youhei Sohma<sup>1,2\*</sup> and Motomu Kanai<sup>1,2\*</sup>

博士研究員 (5年任期)

ERATOグループリーダー (5年任期)

Received 6 September 2014; accepted 16 May 2016;  
published online 27 June 2016

投稿からアクセプトまで約20ヶ月！！

時間を要する生命科学系との融合研究増が背景に

(図 12)

ドボタンがあります。更にスクロールしていくと、もう少し詳細な情報がある画面に行きます(図 15)。ここでポイントになるのが、左上にある「Version2」というものです。プレプリントで未成品ということなので、適宜修正が入って、全部記録されて残るようになっていきます。

左がアブストラクト本文で、右上がどのようなソーシャル・ネットワーク・サービスで見られたかという、Altmetrics という指標で評価されるスタッツです。あとはキーワード、CC ライセンス、エクスポート機能などよくある機能が付いています。

論文の査読をスキップして、専門の研究者、現場の研究者がこういうものを見られるようになったというのは、化学の世界では革命的なことだと思っています。プロの研究者としてどのような使い方をするか考えてみると、一流ラボの論文投稿プロセスが分かるという

ことが非常に勉強になる点です。例えば、こういう題材のものは最初にどういったジャーナルに出すのか、そこにアクセプトされなかったらどういった対策を取っていくのか、それがリビジョンなどを追っていけばある程度分かるようになるということで、私たちからしてみると非常にありがたい情報の宝庫です。

また、査読をする人の感覚も分かります。査読をした人が競争相手だったので理不尽に負けたり、「私の論文を引いていないではないか」と言って、無理やり引用文献を付けさせたりということも実はたまにあるのですが、そういうものもリバイズを見ていけば分かるかもしれないというので、今後は査読側も考えてコメントするようになるのではないかと思います。無茶苦茶な査読も、ひょっとしたらこういうものがあることによって減っていくことが考えられます。

**合成化学トップラボの活用事例**

**スクリプス研究所・Ryan Shenvi研からの報告**  
(V1: 2017/8/18, V2: 8/20公開)

**“10-step Synthesis of 20-nor-Salvinorin A by Dynamic Strategic Bond Analysis.”**  
Roach, J. J.; Sasano, Y.; Schmid, C. L.; Zaidi, S.; Katrich, V.; Stevens, R. C.; Bohn, L. M.; Shenvi, R. A.\*  
[ChemRxiv DOI: 10.26434/chemrxiv.5318188](https://doi.org/10.26434/chemrxiv.5318188)

(図 13)

**ブラウザ画面**

検索窓、投稿用リンク

論文テキスト

ダウンロード・シェア・引用・embed機能

(図 14)

**ブラウザ画面**

10-step Synthesis of 20-nor-Salvinorin A by Dynamic Strategic Bond Analysis

バージョン情報

アブストラクト本文

キーワード

エクスポート機能 (論文管理ソフトなど)

CCライセンス

(図 15)

### 3-3.プレプリント投稿の注意点

図 16 のプレプリントは最初、「files under embargo」という設定になっており、その間は、誰もこの論文の中身にアクセスできないようになっていました。「プレプリントなのに全部隠すの？」と私は思っていたのですが、後々オープンになったのです。

オープンになって論文のフォーマットを見てみると、「Journal of the American Chemical Society」というアメリカ化学会 (ACS) が出しているフラッグシップジャーナルに投稿していたことが分かりました。著者はなぜこういうことをしなくてはいけなかったのかということですが、ジャーナルごとにプレプリントの取り扱いが大きく違うことが原因かと思えます。

図 17 は論文投稿規定です。例えば、「Nature Chemistry」は本当に有名な化学雑誌ですが、この論文投稿規定には、「unrefereed web preprints do not compromise

novelty」とあるので、プレプリントを出しても新奇性は毀損されません。

一方で、「Journal of the American Chemical Society」の論文投稿規定には、「that no portion of this or any other closely related work is under consideration for publication elsewhere in any medium」「publicly accessible preprint Web sites」とあり、プレプリントを出したら投稿済み扱いにするということです。「Journal of the American Chemical Society」に出している論文はアクセスできるような状態にしてあってはいけないという投稿規定があるので、先ほどの著者は非公開にしていたのだろうということです。

論文投稿規定は出版社単位ではなく、ジャーナル単位で違います (図 18)。ACS は「Journal of the American Chemical Society」以外にもジャーナルを出していて、その「ACS Catalysis」は触媒分野なのですが、これに関しては、「プレプリントサーバーに投稿しても大丈夫です、その代わりにプレプリントサーバーに出したということをカバーレーターなどでエディターに知らせてください」という一文が付いています。

ですから、同じ出版社の管轄ジャーナルであっても、扱いが全然違うということがあり得るので、それを頭に入れた上で投稿規定をきちんと読んで、投稿戦略を練らなければいけないというのが研究者視点からの話です。

考えなしにプレプリントを投稿してしまうと痛い目を見かねないという例を挙げます (図 19)。「Nature

**先取権スタンプ目的に使ったハック例？**

---

**Development and application of a highly  $\alpha$ 2,6-selective pseudosialidase**  
31.08.2017, 23:45 by Peter Both, Michel Riese, Christopher J. Gray, Kun Huang, Edward G. Pallister, Iaroslav Kosov, Louis P. Conway, Josef Voglmeir, Sabine L. Fillsch

- このプレプリントは「**Files Under Embargo**」に設定され、投稿直後は内容を閲覧することが全く出来なかった。
- 後日embargo設定が外され、原稿ファイルが公開された。**J. Am. Chem. Soc.**誌に投稿していたことが判明 (後述)。

(図 16)

**論文投稿規定に要注意**

---

論文誌ごとにプレプリントの取扱いは大きく違う！！

**Nature Chemistry IF = 25.9**

Submission to *Nature Chemistry* is taken to imply that there is no significant overlap between the submitted manuscript and any other papers from the same authors under consideration or in press elsewhere. (Abstracts or unrefereed web preprints do not compromise novelty).

---

**J. Am. Chem. Soc. (JACS) IF = 13.8**

Submission of a manuscript to JACS is contingent upon the agreement by all the authors that the reported work has not received prior publication and that no portion of this or any other closely related work is under consideration for publication elsewhere in any medium, including electronic journals, computer databases, and publicly accessible preprint Web sites.

(図 17)

**出版社単位ではなく、ジャーナル単位で違う**

---

**J. Am. Chem. Soc. (JACS) IF = 13.8**

Submission of a manuscript to JACS is contingent upon the agreement by all the authors that the reported work has not received prior publication and that no portion of this or any other closely related work is under consideration for publication elsewhere in any medium, including electronic journals, computer databases, and publicly accessible preprint Web sites.

---

**他のACS系列ジャーナル：ACS Catalysis IF = 10.6**

Posting of submitted manuscripts to preprint servers or databases (e.g. ChemRxiv, ArXiv) does not conflict with ACS Catalysis' prior publication policies, but should be disclosed to the Editor at time of submission.

**Editorにはカバーレーターで知らせるのがベター。**

(図 18)

Chemistry」はプレプリントを出しても大丈夫なので、「Nature Chemistry」に投稿して、早く成果共有もしたいからプレプリントも投稿しました。でも、「Nature Chemistry」には不幸にもリジェクトされました。では別の雑誌にということで「Journal of the American Chemical Society」に投稿しようとしたら、プレプリントを出してしまっているのでは投稿できないということが往々にして起こるのではないのでしょうか。何も知らずにプレプリントを素晴らしいと言って投稿していると、結局損を見るのは自分だということがあり得るので、きちんと考えて使った方がいいという段階にあるのが化学分野の現状だと思います。

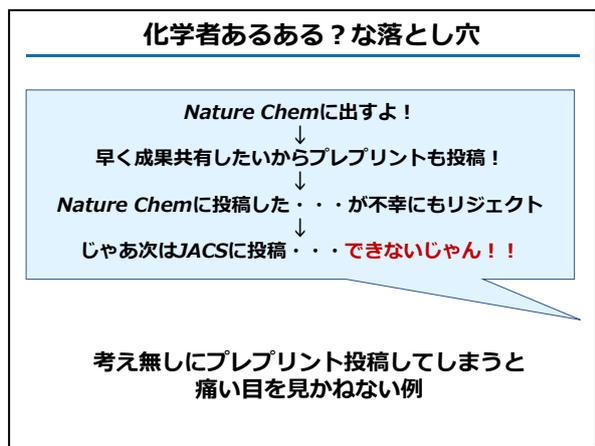
### 3-4.二次情報媒体はプレプリントをどう見ているか

私は Chem-Station をやってきて、サイエンスライター、サイエンスコミュニケーションという視点からものを見るようになったということもあり、実際の論文を分かりやすく紹介する二次情報媒体が、このプレプリントをどう見ているかについてお話しします。

「C&EN」という ACS の二次情報誌は、ChemRxiv に掲載された、論文になっていないプレプリント段階の研究を、2017 年 9 月 1 日に、素晴らしい仕事だと紹介して宣伝しています (図 20)。DOI 識別子も付くので、皆さんここに飛んで見てくださいねということも言えるようになっている仕組みです。

サイエンスライターの感覚からしてみると、プレプリント段階でも宣伝 OK とするスタンスは合理的だと思います。というのも、サイエンスライターは、研究自体の正確性を追究するのが仕事ではなく、研究というものを分かりやすく面白く伝えて、みんなの興味を引くことが仕事だからです。長いとみんな読んでくれないので、興味を引くには、ある程度簡単に文章をつくってお届けしなければいけない。でも、正確性をひたすら突っ込まれると困るということで、みんなそこは頭を使って苦労してやっています。

そういう事情があるので、プレプリントなのだけれど、プレプリント以降のものと比較してどこまで違うのかというと、プレプリント段階でもサイエンスライター的には十分使っていい情報なのではないかという



(図 19)

(図 20)

(図 21)

感覚になるのも無理はないと思います。本当にそれ以上のレベルの厳密さが必要であれば、専門家に監修してもらおうというのが合理的だと思います。

サイエンスライター的に少きついなと思うのが、原著論文（一次情報）へのアクセスができないということです（図 21）。これは本当に大変で、サイエンスライターも正確性に気を使うのですが、正確な情報を得たいと思っても論文に手が届かないということで、やはりオープンアクセスの論文があるといいというのがみんな思っていることではないかと思います。

でも、オープンアクセスジャーナルに出すのは結構大変だということを知ってほしいので、表をつくりました（図 22）。これは化学系の代表的なジャーナルのワールドオープンアクセス化費用です。「Nature Communications」がとてもお高くて約 66 万円です。参考額として、私たちが取れるような若手の科研費の直接

経費は 2 年間で 350 万円です。これに 1 本出すと 1 年間の研究費の 3 分の 1 が吹っ飛ぶということで、事実上これはもう選択肢に入ってきません。でも、これはよくプレスリリースなどになり、良いジャーナルであったりしますから、若手は困ったなということになってしまおうと思います。

この辺のお金の配分も考える余地があると思うのですが、事実上プレプリントで上げて、きちんとした査読システムを通るのであれば、プレプリントでいいのではないかと私たちは思わなくもありません。そろそろ良い査読システムができないかということで、いろいろなところで取り組みをしています。

最近、「Synlett」という有機合成化学のジャーナルが、インテリジェント・クラウド査読というものを始めました（図 23）。100 人ぐらいの査読者にクラウドで議論させて結論を出すというものです。これで結構早く片付くのではないかというアイデアです。レビュアーをきちんと選ぶことが重要らしいです。エディターはきちんと誰が査読しているか分かっているので、査読者を変なことを言わないようにしているということが重要なようです。

最後の結論です（図 24）。今のところプレプリント投稿の勃興・台頭まで来ているので、ジャーナルごとに対応が違っていると困るので、ジャーナルがプレプリント対応の足並みをそろえ、そこからプレプリントに合った査読システムができてくれば、いい感じでわれわれの専門知が世の中に広がっていく世界が出来上がるのではないかと考えています。

**OAジャーナルの研究費圧迫問題**

**化学系フラグシップジャーナルのOA化費用**

ジャーナル	価格
Chem. Sci. (英国王立化学会)	約15万円 (1000ポンド)
Angew. Chem. Int. Ed. (ドイツ・Wiley)	約33万円 (2500ユーロ)
J. Am. Chem. Soc. (アメリカ化学会)	約17万円 (1500ドル, ACS会員限定)
Nature Commun. (Springer-Nature)	約66万円
参考: 科研費 若手 (B)	直接経費 350 万円/2年 (※100%充足率)

参考: <http://orgchemical.seesaa.net/article/43844960.html>

図書館予算をオープンアクセス費へ振り替えても・・・

- ・ ジャーナルを置かない図書館の存在意義とは？
- ・ 科研費 (=税金) が海外出版社に流れ続ける問題
- ・ 金欠の若手研究者は良いOAジャーナルに出せない？

**適した査読システムさえあれば、プレプリントでもいい？**

(図 22)

**インテリジェント・クラウド査読**



Crowd-based peer review can be good and fast

Confidential feedback from many interacting reviewers can help editors make better, quicker decisions, explains Benjamin List.

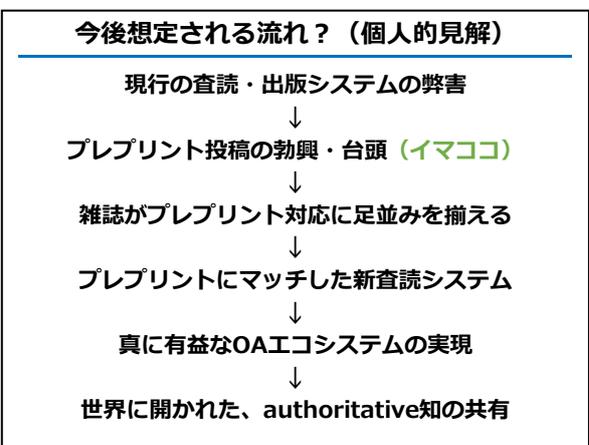
30 May 2017 Nature 2017, 546, 9. (2017/5/30)

**有機合成化学の専門ジャーナルSynlett誌による実験的取り組み**

レビュアーは編集委員会の推薦と自主的に参加する研究者100人で構成され、出版後査読のように、オンライン上で論文にコメントしたり、他のレビュアーのコメントに返信することができる。スピーディーな査読が実現するだけでなく、コメントが匿名で行われるため、査読者によるバイアスのない、率直な意見交換が期待できる。(STI Updates 2017/6/15より)

出版後査読と従来型査読のいいところ取りを狙っている。  
**プレプリント投稿と相性が良い、次世代型査読システム？**

(図 23)



(図 24)

●フロア 1 プレプリントを普及させるという観点でのお考えをお伺いしたいです。bioRxiv にしても ChemRxiv にしても、それなりにサイエンスの分野では受け入れられていると思うのですが、国内の日本語雑誌の話ですが、私自身、日本語の解説記述とかオピニオンのようなものを、編集者とやりとりして、プレプリントとして公開したのも投稿するというので、figshare を使って出したことがあります。その際に編集者の方がものすごく懸念されていたことがあります。プレプリントなので「ここに投稿します」と書くのはいいだろう、でも、国内雑誌で日本語なので、いわゆる研究者の方といいますか、別に研究職に就いていない方も自由に学会員であれば投稿できるというもので、サイエンティフィックには疑問があるものも非常に投稿が多いという問題があります。プレプリントで、査読済み学術論文とは異なることを世間の人は知らないわけです。「この雑誌、この学会のものに投稿中です」と言って出して、それが広く流通してしまう。そして、その投稿中というものを、日本のサイエンライターが一次情報にアクセスしにくいというのは割と話題になっていて、非常に懸念されていたのです。

こういう懸念というのは、理屈上は bioRxiv にしても ChemRxiv にしてもある程度、はねるとはいえあり得ると思うのです。今後、プレプリントが広がる上でそういう問題が発生し得ると思うのですが、そのあたりについてお考えがありましたら聞かせてください。

●生長 クオリティの低い論文が流通して、それを受け止める方が間違っただけで受け止めてしまうことによる弊害というお話だと思うのですが、当然それはあり得るのではないかと思います。研究者のような専門知識を持っている方であれば、恐らくそこを補正しながら使うことができる、別にプレプリントでも普通に査読を

経て出ている論文でも、はなからは信用しないという目線で評価できると思うのですが、では、そうでない方がどうやって使うかということかと思えます。

そこは、それぞれの方に、この情報源はこれぐらいの信頼度を持っているということを理解していただいた上で使っていくしかないものなのではないかと個人的には思っています。ですので、例えばサイエンライターの方でも、一次文献にアクセスできるのであれば、そちらを情報源として重く捉えて文章に書き加えたり、プレプリントであれば、この段階の情報なのだけれどと言外におおむねのような表現を使って書いたりするのが誠実な対応なのではないかと思えます。

●フロア 1 私は化学の分野ではないですが、化学の分野でこの ChemRxiv は始まったばかりとはいえ、プレプリント業界で類似の課題は少なくとも話題になっていないような気がするのです。実際の事例はありますか。

●生長 そもそもプレプリントを専門家以外の人を使うというケースがそこまで多くないのかもしれませんが、また、日本ではまだサイエンライターがそこまで市民権を得ている職業ではないというのが一つ大きな問題かと思えます。これからサイエンライターが稼いで食べていけるような職業になるのであれば、恐らくは長期的に見て絶対出てくる問題かと思えますが、現段階でそこまで問題になっていないのではというのが私の見方です。

●フロア 2 NISTEP の林 (SPARC Japan 運営委員) です。今の議論で思ったのですが、プレプリントにライターが先に価値を付けて記事を書くというのは、査読に加わる、あるいは査読に影響を与え得るのではないかという気がします。ピアレビューが終わる前に「この論文、すごい」と言ってしまったら、それは一つのクオリティアセスメントですよね。そうすると今後ライターがピアレビュー的なものにコミットするような

社会が、日本はまだ構造的な問題でなっていませんが、アメリカなどでは進むような気がするのですが、いかがですか。

●**生長** そういうことはあり得る話かと私個人も思っています。Gordon さんが講演で言っておられたように、サイエンスに取り組む人がフラットになるという思想につながっていくのではないかと思います。

もちろんサイエンスライターだってある程度の専門性を持ってやってきていて、全部に全部はコメントできないかもしれませんが、例えばここに関してはすごく鋭いコメントをするという人がたくさんいていいはずなのです。サイエンスライターの出自のようなものをきっちり管理して反映させるようなシステムがあれば、そういうことはあり得る話かと思えます。

●**フロア 2** 却下されるようなプレプリントを取り上げてしまうと食べていけなくなるので、サイエンスライターのフィルタリングの機能というか、ブランディングにもなるのではないのでしょうか。

●**生長** そうだと思います。また、論文の価値を評価できるサイエンスライターなのと言えるようになる道具にはなるかと思えます。

●**フロア 3** NII の北本と申します。今の質問と関係するのですが、私も昨年、プレプリントサーバーの話で SPARC Japan でしました。やはりプレプリントは玉石混交なので、目利きが必要になると思います。逆にその目利きのブログを購読するようなモデルができるのではないかと考えています。その目利きがサイエンスライターになるのかもしれないのですが、目利きが玉を選ぶというところで何か新しいビジネスモデルができたと思うのですが、そういう構想はありますか。

●**生長** 先ほどの「C&EN」などは本当にそんな感じではないかと思えます。プレプリントの段階で、大々

的に、素晴らしい研究室の誰それが仕事をして出していると宣伝しています。厳密に言うとこれは査読を経っていないので論文としてカウントされないかもしれませんが、ACS としてはそれが合理的だと判断したからそのようにやっているところがあるのだと思います。

素晴らしい若手研究者を前に出すことによって、ACS 全体に益があるから「C&EN」がそれを担っていて、「C&EN」も Ph.D.を持っていて論文の目利きができる編集者がたくさんいる、そういう事情があるからできる話なのではないかと思えます。そういうビジネスモデルを現実的なものにしていこうとすると、おっしゃるとおり目利きができる人材の層をもっと厚くしないとかなり厳しいと思っています。今後どうなっていくかと言われると、特に日本ではどうかというのが個人的な懸念事項ではあります。

●**坊農** 私から一つケムステの方で聞きたいことがあります。私からしたら、大きなコミュニティに育っていると思うのですが、ここまで大きくなった要因とは何なのでしょう。一つは、日本化学会と楽しそうなことをされたりしていることかと思えます。われわれの生物系学会では、「特定の団体と結び付いて」みたいなことを言われて、なかなか実現しないのですが、そちらの方でいい知恵があれば聞かせてください。

●**生長** 当時の日本化学会会長だった玉尾先生がそういうことにすごく頭の柔らかい方だったというのが、一番大きな理由だったのではないのでしょうか。もちろんそれまでにある程度ユーザーやリーダーの数を稼いでいないと、また、絶大な学生からの信頼を勝ち得ておかないとそうはならなかったとは思えます。

それは、私たちは兼業で十年とやってきて、進みが遅いながらもブランディングを重ねてきたというところがあります。もう一つ、そのブランディングに信頼性を与えているものがあるとすると、代表にしろ私にしろ、研究者が本業で、教育にもコミットしないと

いけない人材である大学教員という肩書きを持った人がやっているということだと思います。適当にやっているわけではないというのを言外で感じていただけているところもあるのではないのでしょうか。

●**フロア 2** 化学会にいた人間として言わせてください。なぜかという、代表、副代表とも素晴らしい研究をしているからなのです。2013年に代表の山口先生は日本化学会進歩賞も受賞しています。学会の中でも研究者として非常に優れたお二人だから信頼性がある、広がっているのです。多分、本人は言いづらいと思うので(笑)。

●**坊農** 補足をどうもありがとうございます。そこでちょうど時間になりましたので、これで終わりにしたいと思います。どうもありがとうございました。

## 第2回 SPARC Japan セミナー2017

「プレプリントとオープンアクセス」

# 生命科学分野におけるプレプリントの 位置付けや経験について、統合TVについて

小野 浩雅

(情報・システム研究機構 ライフサイエンス統合データベースセンター)

### 講演要旨



生命科学分野では、2013年にはコールド・スプリング・ハーバー研究所による「bioRxiv」が立ち上がったことを契機に、ここ数年でプレプリントの活用が進んでいる。査読前の論文をあらかじめオープンに共有することで、特に、日々刻々と変化する研究情勢にキャッチアップする必要のある研究者にとって必要不可欠な基盤となりつつあり、周囲の研究者の間では、プレプリントの話題になることが増えてきている。演者自身はプレプリント投稿の体験はまだ無いが、論文投稿の際にあらかじめオープンレポジトリに全てのデータを登録・公開する経験をし、今後このようなプロセスが標準となるのではないかと考えている。



### 小野 浩雅

日本大学大学院生物資源科学研究科博士後期課程単位取得退学。在籍中は哺乳類細胞における脱分化機構の網羅的解析をテーマに研究。2010年より情報・システム研究機構 (ROIS) ライフサイエンス統合データベースセンター (DBCLS) に勤務。生命科学分野の有用なDBの使い方を動画で紹介する「統合TV」の編集や遺伝子発現等の大規模データ解析および可視化、活用支援を行う。2012年6月より現職 (特任助教)。博士 (生物資源科学)。

先ほどの生長さんが化学分野だったので、私は生命科学分野のプレプリントの現状等についてご紹介したいと思います。

そういう意味では、私はもともと wet で仕事をしてきたというバックグラウンドがあります。皆さんもたくさんお持ちだと思いますが、脂を貯める脂肪細胞を

### 1.自己紹介

今はデータベースセンターでコンピュータを相棒にして仕事をしていますが、学部・大学院時代は、細胞などを相手にして顕微鏡を見ながら実験をするというスタイルで研究を進めてきました (図 1)。それを、生物業界だけかもしれませんが、wet 研究といいます。一方、コンピュータをメインに使って生物学の謎を解いていくことを dry 研究とって、最近では wet と dry の融合という話が頻繁に出てきます。

**Wet & Dry**

**学部**  
ブタ成熟脂肪細胞に由来する前駆脂肪細胞 (DFAT-P) の神経細胞への分化転換に関する研究

**修士**  
ブタ成熟脂肪細胞および顆粒膜細胞における脱分化機構の網羅的解析

**博士**  
ブタ成熟脂肪細胞および卵胞顆粒細胞における脱分化ならびに多能性獲得機構の網羅的解析

© Licensed under CC-BY 4.0 © 2017 Hiromasa Ono (DBCLS) Images © 2016 DBCLS Topo Picture Gallery / CC-BY 4.0

(図 1)

他の細胞に変えるという研究をしていました。一番分かりやすい例は山中伸弥先生の iPS 細胞です。あれも結局は分化転換といって、全能性のある他の細胞に変えるという研究でノーベル賞を取られたわけです。その一分野と捉えていただければと思います。

この分化転換、脱分化という現象は不思議な現象なのですが、十数年前にそこを追い掛けるにはコンピュータの力を借りないといけないという状況に直面しました。縁があって、今日のモデレーターである坊農さんのところに、コンピュータで実験する手法を学びに行ったというのが私の最初の出会いです。博士後期課程のときに、コンピュータを使ってこういう生命現象を解いていくということに取り組みました。

そして、データベースを使って生物学を解いていくことを推進するライフサイエンス統合データベースセンターが、ちょうど私が博士の学位を取るぐらいのときに出来上がり、面白そうだなと思いました。周りには優秀な wet の研究者がたくさんいたのですが、こういう研究をする研究者はあまりいないのではないかと、思って飛び込んだのがきっかけです。

私は学生のころからアルバイトとしてライフサイエンス統合データベースセンターに参画していたのですが、当時、ライフサイエンス研究の有用なデータベースやウェブツールが出はじめてきていて、私はそれを動画で紹介する試み「統合 TV」を始めました。自分で動画を編集して、ストーリーをつくって公開していました。

今年の7月か8月で開始から丸10年たち、動画が今1,300本に届こうというぐらい貯まっています。私がアルバイトで始めた当時は、1日行って1本動画をつくるという形でしたが、今はいろいろなお手伝いをしてくださる方のご協力を頂いて、私自身はつくらず、編集者として、きちんとした内容になっているか、分かりやすくなっているかをチェックしたり、ウェブサイト自体を運営したりということをしています。最近言いはじめたのですが、これはまさに生命科学分野の「オープンエデュケーション」ツールではないかと自

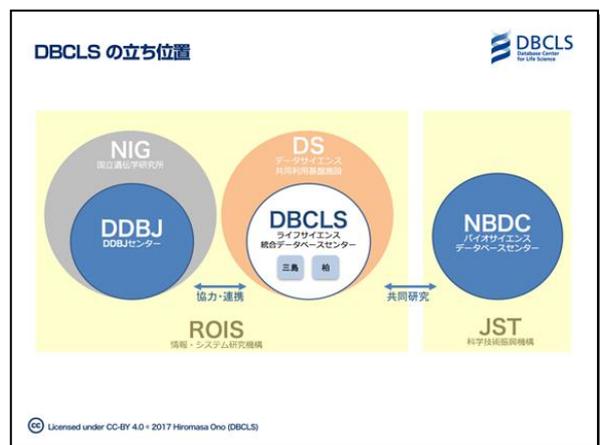
負しています。

もう一つ、生命科学の研究者が実際にどういうデータベースがあると日々の研究活動をより進められるようになるかということを考え、便利なデータベースをつくっていくことも、われわれのセンターの役目の一つです。われわれは実際に実験をしているわけではないのでデータを出すことはできないのですが、既に世界中に便利なオープンデータはたくさんあるので、それを再利用して、例えば、正常組織や細胞株の遺伝子発現データを簡単に検索できるデータベースなどをつくっています。クリックしていくだけで、例えばある遺伝子が心臓でどれぐらい出ているというのが簡単に分かります。このデータベースを RefEx と呼んでいます。

また、生命科学分野のデータベースをどうやって活用していくかというのを、われわれは基本的にはウェブで発信していくのですが、やはりフェース・ツー・フェースで会ってハンズオンにしないと、「これは便利だから使ってみよう」という話にはなかなかならないところがあるので、大学や研究所に出張して、データベースの講習会をよくやっています。そこで講師としてデータベース活用に関する教育もしています。

## 2.DBCLS の立ち位置

ライフサイエンス統合データベースセンターは、長いので DBCLS と呼ばれています (図2)。国立情報学研究所と同様、情報・システム研究機構 (ROIS) の



(図2)

中にあります。似たような研究施設として、国立遺伝学研究所が静岡県三島市にあり、そこと兄弟研究所という形になっています。

DBCLS は三島と柏にラボがあります。DNA を研究者が解析したときは必ずどこかに登録しないとイケないのですが、その日本における元締めが DDBJ という DNA のデータバンクで、遺伝学研究所の中にあります。DDBJ は毎日ものすごい量のデータが集まってくる施設です。われわれはデータベースセンターなので、これらの「ビッグデータ」を活用するために、三島にもラボを構えて DDBJ と密に連携しています。

もう一つ似たような名前のバイオサイエンスデータベースセンター (NBDC) というものがあります。NBDC は科学技術振興機構 (JST) の傘下であって、DBCLS はここと共同研究をしています。

謎の組織 DBCLS をどう紹介するのが適切かと考えたのですが、このセミナーは図書館系の方が今回多いと聞いたので、皆さん FAIR はご存じですよ。あの FAIR が実は、DBCLS の活動の中で出てきたと言ったら、どうですか。DBCLS は筋がいいなと思われた方がおられれば幸いです。見つけやすくて (Findable)、アクセスしやすくて (Accessible)、相互運用できて (Interoperable)、再利用できる (Reusable)、の頭文字を取った標語です。これについて書かれた論文が Scientific Data に出ています。

(<https://www.nature.com/articles/sdata201618>)

このファーストオーサーである Mark Wilkinson さんが DBCLS の片山俊明さんと仲が良く、二人の長年の構想で実現したのが、10 年前から行っている BioHackathon (バイオハッカソン) というイベントです。今年で 10 回目を迎えました。

Hackathon という言葉は最近 IT 系などでよく出てきます。Hack と marathon の造語で、1 週間ぐらい合宿で hacking し続けるのです。ハッキングとは悪い方のハッキングではなくて、コンピュータで仕事することを hack といいます。昼は議論しながら仕事をして、夜はお酒も飲みながら議論する、それを 1 週間ぶっ続

けでやるイベントです。ここで FAIR という概念が大いに議論されたということがこの論文にも書いてあって、Acknowledgements に、NBCD/DBCLS BioHackathon 2015 のおかげでできたというようなことが書いてあります。そういう組織、集団だと思っていただければよいと思います。

### 3.統合 TV

#### 3-1.概要

次に、統合 TV について紹介します。統合 TV は、生命科学分野の使えるデータベースやツールの使い方を動画で手取り足取り解説するウェブサイトで、図 3 のようなインターフェースになっています。生命科学研究を進める上で最低限知っておかなければいけないバイオ系のデータベースを網羅しています。

生命科学分野には、バイオインフォマティクスといって、コンピュータを使って生命科学現象を解き明かしていく分野があります。この分野は、先ほど wet と dry という言葉があると紹介したと思いますが、やりたくない人はやりたくない分野で、黒い画面に緑色の文字が流れてくると思考停止してしまう人がおられるのですが、そういう人に向けて、「実は怖くないのです、簡単にコピーをするだけで緑の画面を克服できるのです」ということも紹介しています。全ての動画に DOI を付けており、引用ができるようになっています。

この動画は YouTube に全て同じものが上がってお

(図 3)

り、おなじみのインターフェースで閲覧することができるようになっています。おかげさまで再生数も右肩上がりで、90万弱まで来ています。

**3-2.講習会の動画**

動画の種類は、単純にスクリーンショットを撮って動画の使い方を紹介するというのが最初だったのですが、われわれはいろいろなところで講習会をしているので、その講習会の様子を動画で撮って、後から振り返れるようにアーカイブしています(図4)。講習会では「今日は何をやります」という形で、説明書や手順書を用意されると思うのですが、それを実際にウェブサイトに乗せておいて、動画を見ながら、手順書を使って、自分の手元のコンピュータで再現できる、再学習できるという用意をしています。このページにはDOIが付いていて、再利用できます。

この分野には最新の解析手法があって、今、ヒトゲノムは非常に安い値段で解析できるようになっています。1,000ドルゲノムといわれているぐらいで、昔は1個解析するのに何百億円とかかっていたのに、今はそれぐらいの値段で、しかも短時間で解析ができるようになりました。その一つのキーワードが次世代シーケンサー(NGS: next generation sequencer)です。これが開発されたのが大きなブレークスルーだったのですが、この解析手法が非常に難しいのです。

生物学はコンピュータを使わないと解析ができないレベルになってきています。その最近の解析手法の講

習会の動画(図5)は、コンピュータリテラシーとサーバ設計の話から始まっており、そういうレベルのデータを扱う必要が生命科学者の研究者の中にも出てきたということを意味しています。後の方になると生物学的な話になってきます。

この講習会は3年前に第1回があって、2014年分は延べ43時間を超えています。43時間もYouTubeの動画をご覧になったことがありますか。1本は2時間半や3時間で、これでも黙っている時間などをカットとしているのです。この講習会は、1週間通しの集中講義形式で東大農学部で行いました。

NGSで職を得ることはできると思いますが、プロとしてやっていこうということであれば、番組を最初から最後まで見てやってみて、やっていけそうかどうかを試してみるといいと紹介しています。毎年、時間数が増えているのですが、今年もあって、絶賛編集中です。

**3-3.生命科学分野の静止画素材**

動画だけではなく、静止画の素材も統合TVで提供しています。もともとはTogo Picture Galleryという名前の別のサービスでした。しかし、ウェブサイトが別だとなかなか普通の人にはたどり着けないというご意見を頂き、それだったら有名な方の統合TVにくっつけようということで、1年ぐらい前に合体しました。

これもわれわれのセンターでコンテンツをどんどんつくっていくために、RAさんを読んで手伝ってもら



(図4)



(図5)

- ・2014年分で延べ43時間超
- ・2015年分で延べ47時間超
- ・2016年分で延べ64時間超
- ・YouTubeのリスト機能で連続再生も可能

いました。ある学生が、動画をつくるのはあまり上手ではなかったのですが、絵を描くのが非常に得意で、「絵でもいいから描いて」という軽い気持ちで始めたのですが、今や500本ぐらい素材がたまっています。これはクレジットを明記するだけで、どのように使っても、商売に使っても構いません。誰でも自由に、閲覧するだけではなく、使うことができる、ライフサイエンス分野の画像やイラストです。種類も生命科学分野に何かしら引っ掛かっていればよいというポリシーでやっているの、各種取りそろえていて、研究発表のスライドや資料などに使っていただけます。よく広告の挿絵に使いたいというお問い合わせも頂いています。

最近では細胞やマウスなどの3Dアニメーションも取り込んでいます。3Dプリンターをお持ちの方はなかなかおられないと思うのですが、最近、街で3Dプリンターを貸し出しているところもありますので、使っていただくと楽しいと思います。

### 3-4. RefEx (Reference Expression Dataset)

冒頭で紹介した RefEx は、正常組織や細胞の遺伝子の発現量を簡単に検索できるサービスです(図6)。遺伝子発現解析の基準となるデータを快適に検索しようというものです。キーワード検索もできますし、「組織特異的に発現する遺伝子を見る」という欄から臓器の絵をクリックして検索することもできます。

2017年8月の終わりぐらいに、この RefEx に関する

論文が出ました。奇しくも FAIR の論文が出たのと同じ Scientific Data というデータジャーナルにです(図7)。これはまさにオープンアクセスです。私自身は実はプレプリントを投稿したことはまだないのですが、データジャーナルに投稿した経験を話したいと思います。

この Scientific Data 自体も生命科学分野で珍しく、普通のジャーナルとは違い、よりデータにフォーカスを当てて紹介するという、チャレンジングな試みです。もともと、研究者が出した、既存のデータベースに載っている一次データセットを他の人が再利用しようとするとき、適切に再利用するための情報が少ないという声がありました。それを一生懸命書いても一次データを出す人にとっては何の業績にもならないからです。これでは Win-Win ではないということで、データジャーナルが出てきたのだと考えています。まさに「一次データセットについて測定の対象、方法、品質を記述する」ために始まったジャーナルだったので。

しかし、われわれは自分自身でデータを出しているわけではなく、それを再解析して使えるウェブツールをつくっているの、この規定とはそぐわないところがあつたのですが、ちょうど運良くタイミングが合い、Scientific Data が再利用を決定的に促進するシステムや技術についての独自のレポートとしてアーティクルタイプの論文を新しく受け付けることになりました。ただ、アーティクルで出すときも、Data Descriptor と同じように、使ったデータについてしっかり記述しなけ

(図6)

(図7)



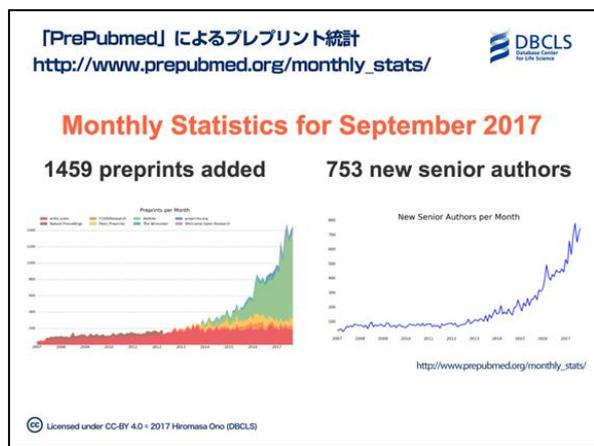
bioRxiv で、面積が急激に増えていることです。bioRxiv が出てきたのを契機にプレプリントがはやってきたということが分かると思います。オーサーも、ちょうど 2014 年ぐらいに増えてきたということが見受けられます。

bioRxiv は 2013 年 11 月にアメリカのコールド・スプリング・ハーバー研究所が開始しました。投稿自体は自由なのですが、論文の体を成していない、内容がふさわしくないというものは選別されているようです。これはどこのプレプリントでもそうだと思いますが、査読後に出版される前の研究成果の迅速な流通とフィードバックの活性化を目的としています。bioRxiv は現在までに 16,000 報のプレプリントが投稿されています。最近のアップデートで、有力な学術誌にそのまま転送して投稿できる機能も付いたそうです。e メールや RSS、Twitter でのアラートシステムが充実してい

るといのが特徴の一つです。

図 13 が生命科学分野のプレプリントのポリシーです。Wikipedia にこのようなエントリーがあって、随時更新されています。どのぐらいのジャーナルでプレプリントを認めるのか調べてみたら、プレプリントを認めないところが意外と非常に少なかったです。これらは独立系のジャーナルで、五つぐらいはまだ認めていないようですが、それ以外は compatible な状況のようです。

DBCLS の同僚の Tazro Ohta (大田達郎) さんは、図 14 のお花見メタゲノムについての論文を bioRxiv に投稿しています。お花見メタゲノムとは、お花見をして、桜の細菌叢をみんなで集めてそのゲノムを読むという面白いプロジェクトです。bioRxiv に載せているだけではなく、ジャーナルに投稿中で今リバイズをしている状況です。



(図 12)

bioRxiv への評価を Twitter などで見ると、非常に好意的な方が多いです。好意的な意見を集めてきている面があるので話半分に聞いてほしいのですが、ある人は「bioRxiv にプレプリントを置いたら世界中から反響が来て、論文はまだアクセプトされていないけれど、別の国際共同研究が始まった」「有名どころが自分たちの材料を使って実験してくれていて楽しい」ということを言っています。

もう一つは、実感できる内容だと思うのですが、「読みたい雑誌の記事が有料だったから読めなかったけれど、タイトルで検索したら bioRxiv に PDF があっ

Journal	Publisher	Policy type	Policy text	Source
Journal of Botany of the Coll.		Likely incompatible	Unclear. When submitting a manuscript to JBC it is understood that the article has been neither published nor submitted for publication elsewhere in any language including abstracts for conferences or less than 100 pages and without restrictions.	JBC/J
New England Journal of Medicine		Incompatible	NEJM expects that the article's publisher will not have been published or pre-accepted before they are submitted to NEJM. The article must have the author's name and affiliation. However, the journal has some other public policies.	JNE/J
Research	Wiley	Incompatible	Based on journal's compatibility with author staff, decision on deposit services is considered prior publication.	Wiley
Journal of Clinical Investigation	ASBMB	Incompatible	The JCI considers the following to be prior publication and therefore unacceptable: <ul style="list-style-type: none"> <li>any printed or online material describing the findings, methods, or results of a submission (as requested in studies of JCI) made available to the public in any form (including preprint servers) prior to the submission of a manuscript to JCI; or</li> <li>any published (online or otherwise) manuscript that in any way duplicates the content of another manuscript submitted to JCI; or</li> <li>any published (online or otherwise) manuscript that in any way duplicates the content of another manuscript distributed other than to a professional audience (i.e., based on personal or internal communications with colleagues or 2016/12/18 Release under Public Communication). We do not consider preprint servers to be prior publication.</li> </ul>	JCI/J
Molecular Microbiology	Springer	Incompatible	FASEB permits the submission of preprint manuscripts, which will change the name from preprint to "preprint" manuscripts. Preprint submissions must meet the following criteria and conditions (1-7).	JMI/J
FASEB Journal	FASEB	Compatible/Compatible	Preprint (Original manuscript submitted to FASEB) can be posted.	JFA/J
The BMJ	BMJ Publishing Group Ltd	Compatible	Submissions will not consider any original research paper or commentary of a research paper that has been published or is under consideration for publication elsewhere. Distribution or preprint servers for consideration prior publication and may compromise the originality of the paper as a submission to BMJ. Authors should not allow posting of preprint papers on non-peer-reviewed preprint servers such as arXiv and bioRxiv. Please contact the editors with questions regarding alternative pathways to other servers.	JBM/J
Science	AAAS	Compatible	Posting to preprint servers such as arXiv or bioRxiv, is permitted. See the PNAS statement on preprint submission for details, and see section 4 for media organizations.	JSC/J
PNAS	NAS	Compatible		JPN/J

(図 13)

(図 14)

ていて読めた、ワーイ」というコメントです。このように bioRxiv を使っている方もいるという状況なのだと思います。

## 5. プレプリントのメリットとデメリット

最後に、生命科学分野に特化した話ではないと思いますが、プレプリントのメリットとデメリットを自分なりにまとめてみました（図 15）。

メリットは、まず、即時性があること、オープンアクセスであること、アイデア・発見の先取権が取れることが挙げられます。また、今はみんながスマートフォンを持ち歩いている時代で、気軽に SNS で情報が行ったり来たりしているので、注目されるというメリットが一つあると思います。さらに、それによって幅広いフィードバックが得られます。

もう一つは、新奇性のないデータを発表できるということだと思います。生命科学に限らず、オリジナリティのない論文はジャーナルベースで世にあまり出ません。再現性を確認しただけの実験や、やってみたのだけかどうかよくいかなかった実験は、既存の媒体に載りづらい性格のものですが、既存のジャーナルのアーティクルの形式にのっとった形で公表できる媒体はメリットだと思います。

デメリットは、不正確な研究成果がたくさん出てしまう懸念があることです。また、コアのアイデアを盗まれて転用されてしまうことを心配する現場の研究者の声もあるように思います。そして、権威がない



(図 15)

というか、今までの既存の評価軸に乗らないところがあるので使いたくないという意見もあるように聞いています。

●フロア 1 NICT の研究者です。生物はプレプリント文化を先進的に切り開いてこられたということで、学会内でどの程度コンセンサスがあるのか、若手は業績に対して安心感を持ってできているのか、研究者の日常の目線から見るとどんな感じかを共有いただけたらありがたいです。

●小野 私が見聞きしている範囲ですが、やはりまだ「プレプリント、ああ、何か聞いたことあるね」というレベルが平均的な生命科学系研究者の感じ方だと思います。今のところはアーリーアダプター、アンテナの感度が良い人が、先ほど紹介した Twitter もそうですが、たくさん発信もして、発信されたものも受け取っていて、「いいね」という雰囲気になっているのだという感じがしています。それが学会としてとか、普通のインフラとなっていくには 2~3 年、あるいはそれ以上かかるかもしれません。

評価されるということに関しては、例えば科研費の申請書には、bioRxiv に挙がっているけれど、「(査読なし)」ときっと書くわけですよ。それを見た評価者が、「何だ、査読なしでは駄目だ」と評価するのか、「プレプリントを使っているとはなかなか筋があるな」と捉えるかどうかという、評価者個人がどう思っているかにまだ依存するところがあると思います。そこでどのようにコンセンサスを得ていくかは、生命科学に限らず、学会なり、国なりがどういう施策をしていくかということに依存するのではないかと思います。

●フロア 2 NBDC の職員です。小野さんがご存じかどうか分からないのですが、「Nature」の記事だったと思いますが、bioRxiv に載っている論文は、厳しめ

のライセンスを付けている人が多いということが載っていました。クリエイティブ・コモンズ (CC) を付けることもできるのですが、そもそも CC を付けずに普通の権利を付けていたり、CC でも特に厳しい改変禁止を付けていたりするものが非常に多いという話があります。

一方では、どうもその記事を読んだ印象では、bioRxiv のポリシーでは論文に関してはテキストマイニングが全てフリーで行えるということなのに、大半の著者はそのことを意識せずに投稿しているのではと思うのですが、その辺について聞いたことはありますか。

●小野 私もその話を聞いて、「へー」と思ったところですよ。やはりまだ bioRxiv についても、行き届いていないという批判的な記事も見つけました。プラットフォームとして今一番有名なのでみんなこれを使っていますが、まだ知らないでライセンスを付与しているところがあるので、その辺をきちんと意識してライセンスを付ける、付けないという意思表示がきちんとできるように、まだまだプラットフォームとしての改善の余地があるのではないかと思います。

●フロア 3 NII の北本と申します。データジャーナルへの投稿で解析済みデータを figshare に寄託したということについて、例えば DBCLS そのもので公開することはできないのかということと、日本のデータが figshare に流れてしまうことについて、こうしたらいいのではないかと意見ををお願いします。

●小野 非常に重要なご意見ありがとうございます。われわれは公開リポジトリに出したのですが、Scientific Data が認める公開リポジトリであればどれでもいいという形になっています。NBDC のデータベースアーカイブというサービスも実は Scientific Data が認める公開リポジトリなのです。しかし、これは言っているのか分からないのですが、やはり figshare は楽

なのです (笑)。これはインターフェース的な問題です。われわれも NBDC と共同研究をしているので、日本のところに置くべきだろうというご意見を頂くのはごもっともなのですが、やはりいざ論文を自分で出そうとすると便利な方に流れてしまったというところがあります。

ただ、「figshare に流れてしまう」という表現をされていましたが、彼らが困ってしまうわけでもなし、プラットフォームとして用意されているので、個人的にはそれでいいのかなという気持ちです。NBDC さんにはその辺を頑張っていただきたいという思いです (笑)。

●坊農 補足します。この責任著者として言わせていただきますが、これを始めたときは、NBDC でそういう指定はなかったのです。だから、どこかに上げないといけなかったので、Dryad か、figshare に上げるかと調べたら、Dryad は有料、figshare は無料ということで、お金がないので figshare になりました。

●フロア 4 NISTEP の林 (SPARC Japan 運営委員) です。GitHub の例を出されたように、オープンソース文化が特に dry なバイオサイエンス分野にあるので、私は何となく FAIR 原則はもっと文化として取り込まれているイメージを持ちがちなのですが、それでもプレプリント文化のようなものはまだ進まないのですか。情報系バイオサイエンティストと wet 系バイオサイエンティストの差のように単純に分けていいのか、もう少し複雑な研究者文化背景があるのか、教えていただけますか。

●小野 まさにご指摘のとおり、バイオインフォマテイクスといわれている人たちは、こういう FAIR のようなものは意識していたか、していないかは別として、頭の中にはあった概念だと思います。

●フロア 4 GNU ライセンスなどはもともとからそうで

すよね。

●小野 そうですね。もともとそうなので、むしろ何でオープンではないのか、というところです。私たちの周りでも、バイオインフォマティクスの人たちがかんでいるものが積極的に bioRxiv などにも投稿されています。当然、存在も知っているし、「こういうのに出していこうよ」「今、これが一番受けているんだよ」という感じでプッシュしているという印象があります。ただ純粹に、実験系のラボ、医学系のラボなどだと今のところ「何これ」という話になっていると思うのです。当然そういうことを教えてくれる人もいません。

●フロア 4 所属する組織の文化に引きずられるのですか。

●小野 そうだと思います。周りの付き合いしている人の文化は多分にあります。おっしゃったように、IT系というか、インフォマティクス系は全然これに対する抵抗はないのです。むしろ「何で出さないの」と言われるぐらいです。ただ、それが全部の分野に適用できて、公開前から出していいのかという話にも当然なるので、これが大正義というわけではないと思います。そこは純然たる深い溝がまだあると感じています。

## 第2回 SPARC Japan セミナー2017

「プレプリントとオープンアクセス」

# 全 体 議 論



<b>坊農 秀雅</b>	(情報・システム研究機構 ライフサイエンス統合データベースセンター)
<b>引原 隆士</b>	(京都大学図書館機構長/arXiv.org 会員コンソーシアム代表)
<b>生長幸之助</b>	(東京大学大学院薬学系研究科/化学ポータルサイト Chem-Station 副代表)
<b>小野 浩雅</b>	(情報・システム研究機構 ライフサイエンス統合データベースセンター)

●**坊農** 全体議論を始めたいと思います。質問し足りなかった方はぜひお話ししてください。多岐にわたってお話しいただいたので、個別の先生へのお話などもあると思います。

●**フロア 1** 産業衛生学会のもので、私はジャーナルの編集をしているのですが、海外の医学系ジャーナルでのプレプリントサーバについて、どれくらい広まっているのかなど、状況をどなたかご存じですか。以前担当していた医学系のジャーナルで、中国からの投稿にクロスチェックをかけたところ全部一緒になって、結局プレプリントサーバに入っていたということがありました。そのジャーナルはプレプリントサーバへの投稿を認めないとしていたのですが、小野先生の資料を見ると BMJ は認めているし、以前に Elsevier の方から Elsevier ではほとんどが認めていると聞いたことがあります。海外では実際にはどんな感じかお分かりになりますか。

●**小野** 私も医学系ではないので残念ながら具体的な話は分からないのですが、Wikipedia のリストを見ると、incompatible なものは割と医学系、独立系の雑誌が多かったです。医学系が悪いわけではないのですが、保守的な感じがあるのかなど。先ほどバイオインフォマティクスの方が先に進んでいるという話がありましたが、そういう意味では、中身のセンシティブさに時

代が追い付いてきていないところがあるのではないかとこの印象を持っています。

●**フロア 2** 慶應義塾大学の市古 (SPARC Japan 運営委員) と申します。アメリカ国立衛生研究所 (NIH) がプレプリントに対する方針を今年出したことによって、PrePubMed という動きになってきたのだと私は理解しています。今、保守的だという話があったので少し付け加えますが、確かに慶應義塾大学の中でどの程度 bioRxiv に投稿されているかを見たところ、臨床医学の積極的な投稿は目立たなくて、理工と医学系に関わる方、それから SFC で情報科学に関わる方と医学研究を一緒にやっている方の投稿が見受けられました。

●**坊農** 私自身も少し医学系をやっているのと、BMC Genomics でアソシエイトエディターをしていたこともあるので、それで見ていると、プレプリントに出したものは、リファレンスとして挙げられることはよくあって、普通の論文のように扱えるという反面、基本的には査読を経っていないので信用できないという扱いになっていたと思います。

ただ、小野さんの出してくれた Wikipedia の表も動的に変わっていったら、incompatible だったジャーナル誌が急に compatible になっていくという傾向があります。ざっくり探すときの参考情報としては Wikipedia はいいと思いますが、情報のソースが最新ではないの

で、実際のジャーナルの判断をそれぞれのところへ見に行き、正しいかどうかをよく見た方がいいと思います。ただ、劇的に変わっていている現状があると思います。

医学系で全く同じものを出してくるというのは、論文の本数を稼ごうとしている人がやりがちで、そういうものは情報技術で対抗して取り締まるのが大事だと、私自身やっています。それをやり過ぎで逆に、前の論文と同じ手法を使った論文がはじかれたということが最近ありました。自分が前に出したものとほとんど同じということでプログラムのチェックではじかれてしまって、他のジャーナルに出さざるを得ないという事態になったこともあります。

プレプリントで出したものと全く同じなど、おかしいものはおかしいとチェックできるようになってきたという点では、良い時代になってきたのだと思います。

●フロア 3 国立極地研究所の図書館員です。引原先生にお伺いします。コーネル大学で、arXiv.org のエフォート管理をしながら、ライブラリの中で実際に回しているということですが、arXiv.org は今後、拡大路線で行こうとしていて受け取りました。そのときに、ライブラリで今後どのような運用体制を取っていくのでしょうか。また、その視点から、SSRN では商業ベースに移行したという話がありましたが、そちらの可能性についてはどう考えられているのでしょうか。

●引原 それは非常に微妙なところで、Gordon さんの講演で言及のあった、SSRN が移行した理由を Twitter で拝見したのですが、確かに経営的な問題とか、人のお金の問題は大きいです。arXiv.org の場合はタスクが増えていくのに合わせて、エフォートがあるときは増やしたり、あるときは減らしたりしています。今、従来のシステムから新しいシステムに移るときに増員しています。ファンディングに対して増員を要求してお金を付けてもらうという形を取っていて、それは経常的ではありません。いずれそれは減るので、

その分が今度はチェックする人たちの雇用の方に回るのであるというのが、経営状態を見ていても分かりません。

かなりの部分はオートマチックになっていますが、iThenticate に引っ掛かったものなどのケアは丁寧にやっているので、それが増えてきたときにどうなるかというのは確かに問題かもしれませんが、現状は回っていると思います。

●フロア 4 NISTEP の林 (SPARC Japan 運営委員) です。先ほど引原先生にしなかった質問を投げたいと思います。リソースシェアという意味で、機関リポジトリがせつかく 744、今はもっと増えているはずですが、あるので、アーカイブの受け先を日本の場合は機関リポジトリにしようというアイデアはどうなのでしょう。つまり、会費を払うだけではなく、ストレージの置き場やナビゲーションのところにコントリビューションで組む、商業主義的なものへの対抗として考えられるアイデアとしてそういうものが思い浮かんだのですが。個人的な感触で構わないのですが、そういう方向性をまず示さなければいけないのではないかと、ということが 1 点と、機関リポジトリをそもそも具体的に使うことに関してはいかがでしょうか。

●引原 今年の arXiv.org の Member Advisory Board (MAB) で、われわれ日本から、機関リポジトリについてディスカッションしてほしいと要求を出しました。会議の日の昼休み (13~14 時半) に昼食を取りながらディスカッションしたところ、リポジトリ自体についての考え方がヨーロッパと日本とアメリカでは違うということで、日本の場合はリポジトリがプレプリントサーパ的なところがあるのでその可能性はかなりあるのですが、やはりなじまないという議論がありました。

それは機関がどこまで全体のデータベースのサポートにタスクを出せるかという問題があると思うのです。今はコントロールできるミラーサイトと Amazon のク

クラウドでやりたいというのが正直なところで、京都大学の基礎物理学研究所にあったサーバがなくなったのもそれが一つの理由です。今後はクラウドになっているときに、日本でクラウドをどこかが管理するということは当然あると思います。けれども、現状のシステム開発をしている段階では、やはり全体をマネジメントしたいという方が強いようです。

●フロア 2 少し下世話な話なのですが、arXiv.org は経営的には問題ないのでしょうか。慶應義塾大学の例なのですが、arXiv.org に協力するかどうかというお金の分担の話が来て、誰がどう払うかとなったときに、図書館員が価値を示していくのがなかなか難しい状況にあります。私としては arXiv.org に協力することはいいことだろうとは考えているのですが、お金の面が大丈夫であれば学内でもう少しいろいろ議論できます。将来的にこれが成り立たなくなることがあるとすれば、図書館としては何をどうすることができるのかということをお話いただければと思います。

●引原 今日はお金の議論は下世話になると思って外していたのです。ここ 2~3 年は開発の経費のこともあって増えていますが、それはいずれ落ち着きます。ただ、2018~2022 年のモデルとして、どれだけのユーザー負担を考えるかを議論した結果、利用率に応じた負担を頼むということで今まで来ています。

現状では日本に負担が来ていると思えるのですが、かなり考慮を要求しました。日本の場合はファンドを直接アメリカに送ることは、内閣府や文科省に聞いてみてもやはり難しいということがあって、国立情報学研究所 (NII) などが取りまとめてコンソーシアムで払うという形態しかないだろうというのが向こうとの認識にありました。

それは日本だけではなくて、アジアの多くの国が多分そうだろうということを言っています。もし仮に日本のファンドを、S が付いている会社などに出していただいたら、これは大きな違いがあるかもしれません。

私の認識としては、経費を出すということと同時に、今まで arXiv.org の単なるユーザーであった日本が発言権を持って、いろいろな提案をしてきています。昨年からはかなりして、それで存在感を示すことができています。arXiv.org を中心として、先々プレプリントサーバが統合していく動きになると思うのです。そのときに発言権をかなりキープできるのではないかと考えているので、ここは単にユーザーとしてのお金ということだけではなくて、次のものを得るための投資として考えた方がいいのではないかと考えています。

ですから余計に、例えば「ビューワーをもっとこうしてください」と発言したっていいですし、「メタデータもこういうのはどうですか」と提案してもいいわけです。今までは単なるユーザーだったからできなかったもので、やはりそういう方向に転換した方がいいのではないかと私は思っています。

●フロア 5 先端医療振興財団のものです。私は医療系に携わっていて、各学会との連携もあるのですが、学会同士のデータシェアリングなどの課題があります。ただ、それぞれの学会でいろいろな考え方があるので難しいという状況だと思います。それで、日本で独自にそういうデータベースを作成されるお考えがあるかどうか、それをハーモナイズして海外に展開する構想はあるか、もしないとすればどういうところが難しいか、どの先生でも結構ですので教えていただけたらと思います。

●坊農 私たちライフサイエンス統合データベースセンターは、予算規模的に基本的にはライフサイエンス分野で多くの方に使ってもらえるデータベースの維持・管理ぐらいしかできません。ある分野のデータベースをつくることに関しては基本、関与できないのが現状です。なぜかという、お金というよりも、まず人が足りないからです。そういうことをやったらいいのにと皆さんから伺い、確かにそのとおりだと思うのですが、分野オリエンテッドなデータベースはその分

野でやりたい方に始めてもらって頑張ってもらって、困ったことがあれば、われわれが技術的にお手伝いするというのが現状だと思います。

●フロア 5 それをもし商業ベースでどこか出版社にお願いするとしたら、非常に高額になって、実現が不可能になってきます。今ある海外のものを使うということに関して、それぞれの学会の先生の考え方があって、ここは使うがここは使わないということで、結局ばらばらのデータベースに入ってしまった、それらが結び付かないことが今後問題になってくると思うのですが、そのあたりはいかがでしょうか。

●引原 データベースに関して、海外の出版社の狙いはまずデータを取得することで、次はクロス的に AI などを使ってマイニングしていくことなので、日本の学会がこぞって海外のデータベースにデータを置くことは非常に危険です。ジャーナルで負けたところを、最後データで維持しているのですが、それがぎりぎりの状態にあるということを学会の先生方に理解していただかないといけないのかもしれないかもしれません。

それをどうしたらいいかというときに、クラウドでも、日本の NII を中心とした形でもいいですから、機関がデータサーバをパブリックにサポートするシステムを立ち上げなければいけないのは明らかだと思います。それをやらずに、学会が学会の維持のためにやるということは危険ではないかと思われまます。アーカイブが商業出版に吸収されてしまったという例がありますが、それはアーカイブを維持するためという非常に矮小な議論の結果だと思います。そこは学会が、データのユーザーとしての研究者や企業などと協議して決めていかなければいけないのではないかと考えていただきます。

●坊農 NBDC から、公開データベースではないものの話はないですか。紹介していただけないですか。

●フロア 6 NBDC で提供しているものとしては、私が担当している生命科学系データベースアーカイブというものがあります。これは基本的には全部公開していて、あまり突っ込んだ加工はできないのですが、ある程度再利用できるような形に持っていくことはできると思います。それとは別に、エンバーゴのようなものですが、将来公開することを前提として、プロジェクト期間内はプロジェクトの中だけで共有するというサービスも展開しているので、個別に相談していただければお力になれるかと思います。ただ、基本的には、データをこういうふうにしたいというのはそれぞれ研究コミュニティの方に主導していただきたいと思いません。

●フロア 7 これはデータの保存ということであれば、また別の面があります。2017 年 11 月 6 日 (月)、国会図書館と研究データ利活用協議会 (RDUF) で、データサイエンスを巡る世界の動向におけるデータのアーカイブといった問題について議論するので、そこで続きの議論ができるかと思います。

(<https://japanlinkcenter.org/rduf/events/index.html#s001>)

活用するデータにはあまり適切ではありませんが、データを保存するだけであれば国会図書館も使えるかもしれません。要するに活用するデータはそういうのはあまり適切ではないと思うのです。データとしてとにかく保存したいというのであれば、またアーカイブという視点では別の策があるのではないかと思います。

●フロア 4 今日しなければいけない大事な議論は、どうやってプレプリントが各研究者コミュニティに受け入れられるかということです。なぜ最初に物理分野でうまくいったのか、そのうまくいった要素は他に展開できるのか。私の分野の化学で、生長さんの話を聞いて一つ作業仮説を立てるならば、そのコミュニティのトップジャーナルのエディターが受容するかどうかということがあり得る気がします。そのあたりはパネルの皆さんの領域でどうですか。トップジャーナル

のエディターが「プレプリント、オーケー」と言うと、傘下のジャーナルみんな横並びになりそうな気がして、それが物理で起きていて、化学ではまだ起きていないということだと思うのですが、いかがでしょうか。

●**生長** 全くおっしゃるとおりだと私も思います。端的に言うと、日本化学会がプレプリント投稿を受け付けて、化学会全体でこの流れを進めようと言って、評価基準をみんなで修正していこうという流れをつくれれば、時間がかかるかもしれませんが、徐々にはそうなっていくのではないかと思います。その過程で、プレプリントは有益なものであるということがコミュニティに受け入れられていけばそれで問題はないかと思います。でも、誰がそれをやるのだというのが難しいところだとみえています。

●**フロア 4** お隣に物理学会の関係者もいらっしやいますよ。

●**フロア 8** 物理の中でも、先頭を走ったのはやはり高エネルギー物理なのです。ビッグサイエンスで、データは共有物という意識もあったと思いますし、何とんでもコミュニティとしては小さいので、そこでは合意が得やすかったということがあります。次第に物性物理にも広がっていったのですが、私の見るところ、その契機は、高温超伝導ではなかったかと思うのです。あのときに研究のスピードが爆発的で、査読付きの論文なんか待ってられないと、みんなプレプリントで勝負したのです。それで一気に広がったというのが私の個人的な印象です。

●**フロア 4** これはNIIの北本さんが以前ディープラーニングで紹介した例とまさに同じです。研究が早過ぎて待ってられない。

●**坊農** 北本さん、何かコメントはありますか。

●**フロア 9** NIIの北本です。昨年の第3回SPARC Japan セミナーで話したのは、ディープラーニングで研究の速度が爆発的に速くなって、論文の査読を待てられないからみんなプレプリントに流れたということです。高温超伝導と全く同じというのは今日初めて知ったのですが、それは非常に大きな要因かと思います。

●**坊農** どうもありがとうございます。昨年出ていない人にもお分かりいただけたかと思います。

今の、化学ではこれから起こるとか、そういう話を聞いていて、われわれ生命科学の分野はそういう意味では不思議な立場だと思いました。ビッグラボなどが何かやってというところが日本ではあまり起きていなくて、海外で分野の最先端を行っているところが最新の機器を使ったデータなどをプレプリントで出しているということは、たまに聞きます。日本では生命科学関係ではそういうことがあまり起きていない感じがします。化学などは、まだ学会がきちんとオーガナイズする力を持っているのではないかと思います。やはり日本化学会に頑張ってもらって、プレプリントをやりましょうと言うのが一番いいのでしょうか。

●**生長** プレプリントの普及の順番として、先に物理があって、次に生物が来て、化学に行くという順番は分からなくもないです。実験に時間がかかり過ぎて、査読の速度とキャリアプランに齟齬が起きやすいのが生物学の世界だと思っています。生物学の研究を少ししているので、最近何となく感覚が分かるようになってきました。先ほど言われていた高温超伝導の話は、研究のスピードの方があまりに速過ぎてということですが、評価やキャリアパスなどの人間的なところとサイエンスのスピードのギャップが出てきた瞬間に、プレプリントが普及していくのではないかと思います。

化学の世界では、今までは評価やキャリアパスなどのサイクルと研究サイクルがちょうど良くかみ合っていたので、プレプリントの必要性がひょっとしたらな

かったのかもしれませんが。けれども、化学の世界でも生物との融合的な研究をかなりたくさんの方がやるようになった結果として、プレプリントが必要という話になったという気がしています。それをトップの方が理解してそれをやろうとするかなとは思っています。

●フロア 7 北本さんの補足ですが、コンピュータサイエンスのプレプリントが普及したのには、もともとジャーナルに権威がないという事情もあります。カンファレンスがベースで、カンファレンスにも権威があるもの、ないものがあって雑多な世界です。しかも昔からテクニカルレポートも大学ごとに出すという文化もあって、そういうことに関してもともと無頓着なところがあったのです。今までも、例えばページランクの詳細な論文はスタンフォードのテクニカルレポートしかないということがあったぐらいで、それがたまたまアーカイブに載るようになったのが割と最近です。コンピュータサイエンスの場合は逆にジャーナルにさほど権威がなかったのでそういう方向に流れてもそんなに摩擦はなかったと私は理解しています。

●坊農 確かに考えてみると、私が生命科学でプレプリントでと例に出したのも次世代シーケンサーで、より長く早く安く読もうという競争があって、早く論文を出したがついている分野です。小野さんに出していただいた例も、シングルセル、1細胞の技術という、今競争の非常に激しいところです。生命科学の bioRxiv を見ていると、そういうところしか出ていないような印象を受けています。厳しい競争がプレプリントを促すというまとめでよろしいのでしょうか。そういう競争が化学でもいろいろありそうですが、化学だと次は何がいけそうですか。

●生長 私が属している有機合成は研究サイクルが短い分野なので、タッチの差は当然あるのですが、プレプリントで評価される、されないというほどの差ではない気がしています。どちらかというと化学はもっと

研究スパンの長いもので評価が決まってくるようなものが増えてきたと思います。

●坊農 それでは時間となりましたので、全体議論を終了いたします。先生方、どうもありがとうございました。