

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 のアドバイザリーボードに、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

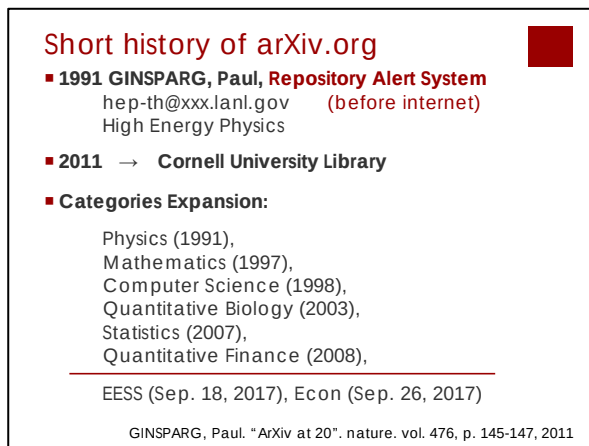
arXiv.org の創始者は Paul Ginsparg で、ロスアラモス国立研究所の彼のデスク上の HP のサーバから arXiv.org は始まっています。それが 1991 年で、インターネットが始まる前です。私が在外研究に行ったのが 1993 年で、その直前は BITNET などが連絡手段でした。その時代にメールベースで始められたのがそもそもその始まりです。

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。

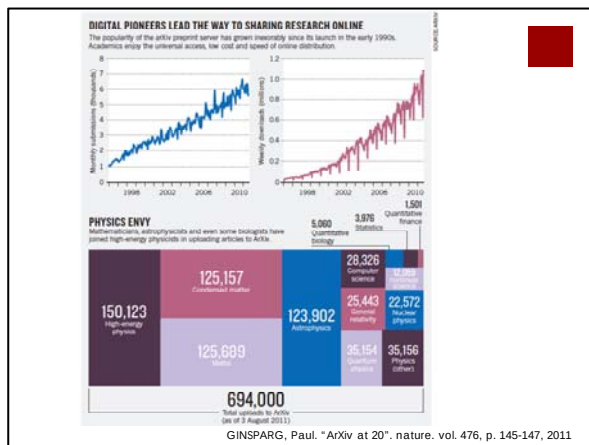
それから 20 年たったことをきっかけに、2011 年にコーネル大学の図書館に移行しました。私自身が在外

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。



(図 1)



(図 2)

さらに、economics が 1 週間後の 26 日に加わっています。

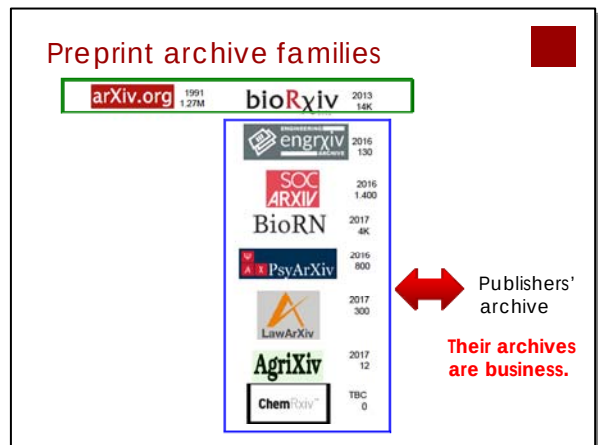
物理で始まって、物理の関連で次に数学が始まったというのは分かりやすいのですが、この後どのように発展していくかということを決めて、現在に至っているかというのが重要なキーになります。

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も情報交換しており、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。



(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点で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

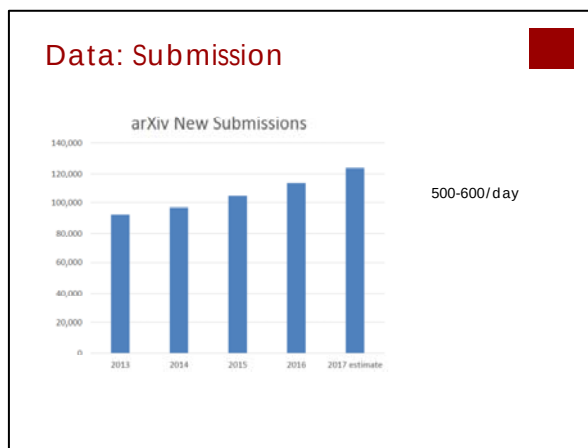
arXiv.org のサブミッション数は 2012 年に約 84,000 で、それからどんどん増えていって、2013 年に約 92,500、2016 年に約 113,380 です。ダウンロード数は 2016 年に 139 百万です（図 4）。ですから、膨大な数のダウンロードが始まっているのです。

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

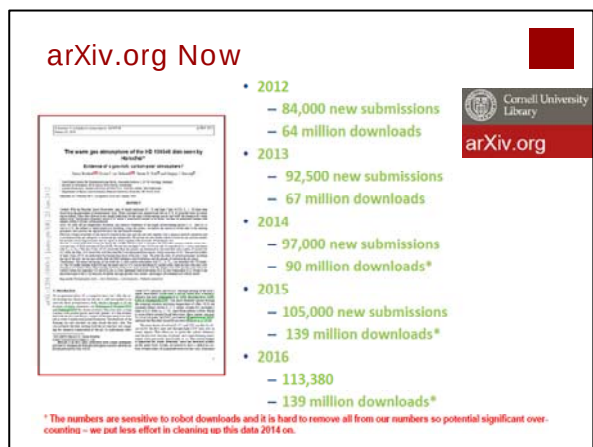
の、それからインターネットが始まったときにこのサイトにアクセスしたら見られたというものから、検索エンジンを介して提供するという状態にだんだん移っています。

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。

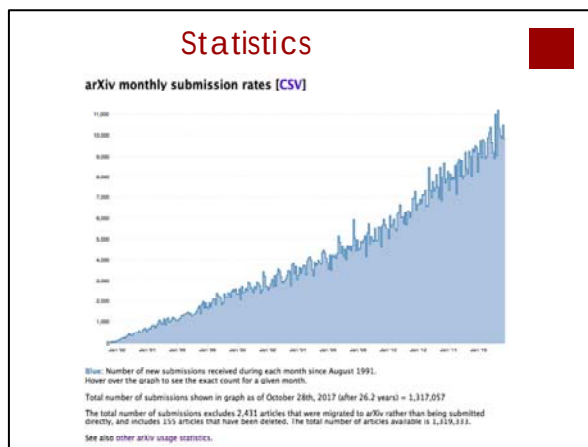
それをもう少し統計的に見ると図 6 のようになっています。1992 年は微々たるものだったのですが、現在はここまであります。サブミッションは、日本とは違い、夏休みはやはり少なく、クリスマスも少ないです。この波が大体毎年同じような傾向をたどっています。これは増加を総量で書いているので、その年々で出すと大体フラットな変化をしています。加算してリニアに増えています。



(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比べてみます(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がほとんどだったのですが、それが時間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。

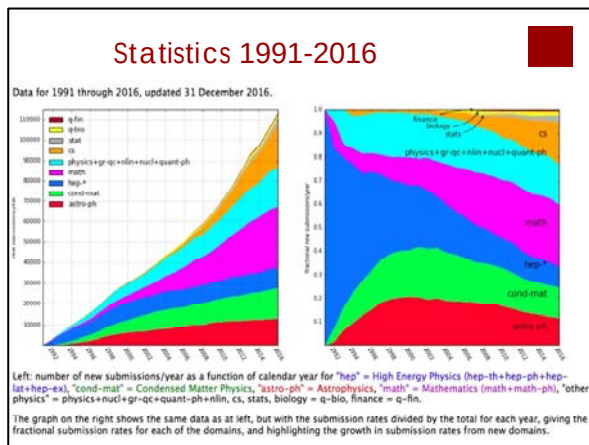
このうち、サブミッションとダウンロードを行っている国の内訳は図 8 です。まず、arXiv.org を支えているアメリカのコンソーシアムが 18%、アメリカのノンコンソーシアムが 21%と結構多いです。一つ一つの大学や研究所がコンソーシアムとして組んでいる部分と、そうでない部分が大体同じ比率になっています。UK が 7%、日本が 7%と非常に多いです。この中でも研究大学といわれるところが多く論文をサブミッ

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。

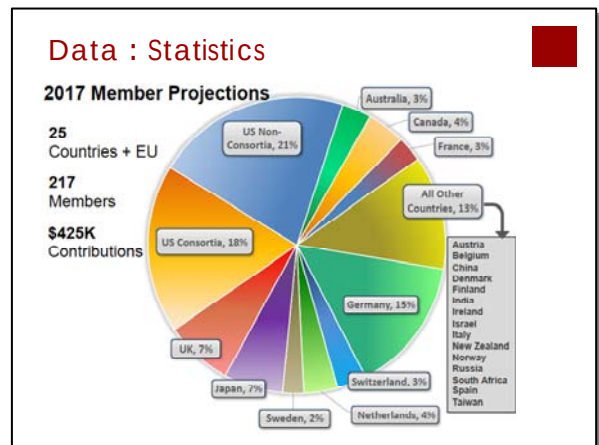
アジアでも増えており、中国のユーザーの割合が非常に高くなっています。また、投稿よりもダウンロードの数が非常に多いということが知られています。このような割合になっています。

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も、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、これ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いけないのです。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、その検索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検索以外に、Google の検索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

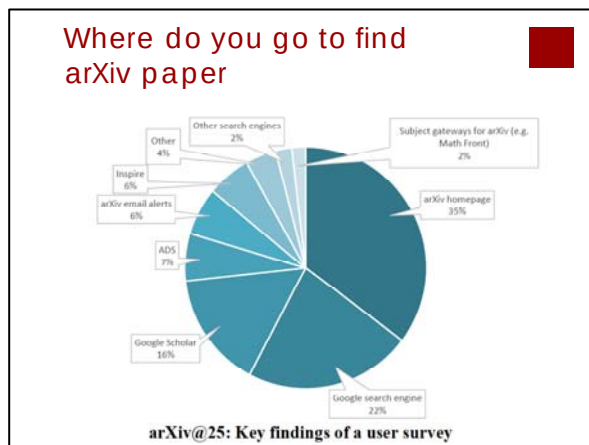
の論文が引用されました」というようなアラームがあると思うのですが、それが結構有効だということがよく分かります。

INSPIRE は別のデータベース検索になりますが、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検索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。

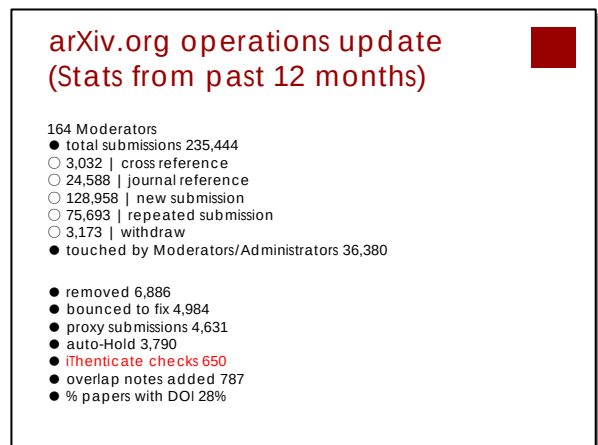
少し中身に入ります。先ほどトータルサブミッションの話をしました。そのトータルのサブミッションは約 230,000 あるのですが、そのうち、ジャーナルのリファレンスが約 25,000、クロスリファレンスが約 3,000、ニューサブミッションが約 130,000、再投稿が約 75,000、withdraw (取りやめ) が幾つかあります(図10)。

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います。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きているということです。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、何年ものという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。

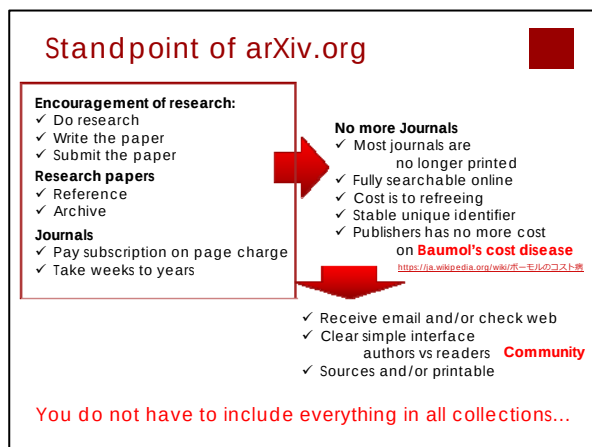
そういうものに対して、出したということをきちんと arXiv.org でオープンにして、それが伝わるようにして、物理なら物理の世界が早く進むようにエンカレッジしていくというのが基本的な考え方です。ですから、やはり物理特有の状況に根差してこうなってきたというのがよく分かると思います。

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、「ボーマルのコスト病」といって、もはやもうける仕組みにはなっていません。「ボーマルのコスト病」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)

ことです。

arXiv.org は今、次世代に向かって動いています。図 12 にはステークホルダーとデベロッパーと書いてあります。そのステークホルダーはリーダーであり、サブミッションするオーナーであったりするわけですが、それ以外にパブリッシャーも当然います。デベロッパーというのは、このシステムをつくっている人たちも含まれているのですが、その中で、どういうシステムを次の世代でつくっていったらいいかという議論を非常に熱心に行っています。これが恐らく次の世代をつくります。

従来のシステムは何でもかんでも入れ込んでいるデータベースでした。これはもう今の時代では成立しません。ですので、それを書き換えて、現在は application programming interface (API) を使って、データベースと API をフィルタリングするような形を取

って、最終的には図 13 のような形を目指します。ブラウザとライブラリがあって、API が管理する business logic と service integration のプログラムを置いておいて、それでデータベースとファイルシステムをやりとりするというようなことです。これらに対して従来型のプログラムをそのまま延長するのではなく、Python や Flask web プログラム、Docker や REST APIs など、最近使われているものをフルに取り入れてセキュリティを保ちながらやっていくということを計画して進めています (図 14)。API に持っていくことによって何が良くなるかというと、今までのデータベースだけではなく、他のものも取り込める可能性が出てきます。そのために、将来必要なイメージや raw data といったものへのデータベースのリンクも可能になるであろうことが予想されます。

実際の絵は図 15 のようになります。arXiv.org のシ

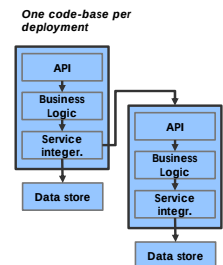
What is the arXiv.org, what is it for?

- **Description at multiple levels:** system context, subsystems, components.
- **Audience:** stakeholders and developers.
- **Not a requirements analysis.** Describes decisions and their rationale, and the most important requirements of the system as a whole, but allows for agility and changing requirements throughout the project.
- **Both prescriptive and descriptive:** commemorates technical decisions in context, provides guidance for implementation, but also evolves as new decisions are made throughout the development process.

(図 12)

Where is arXiv.org going?

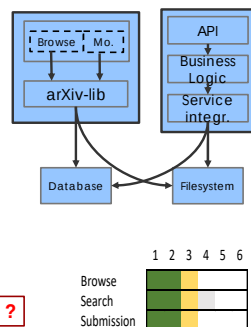
- Fine-grained **isolation with services:** **Python 3 + Flask web** micro-framework, **Docker** containerization.
- Consistently-applied internal architecture.
- Integration via **REST APIs**, notification broker.
- **Polyglot persistence:** isolated data store, choice of technology matches service requirements.
- Independent scaling, server resources reflect demand.



(図 14)

How does arXiv.org get there?

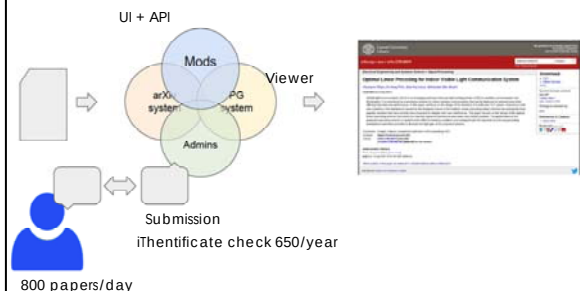
1. **Prioritization:** from the "outside" in.
2. **Identify minimum integrations:** database, filesystem.
3. **Re-engineering:** preserve behavior, but with re-architected codebase.
4. **Local deployment:** services can be deployed on existing web servers.
5. **API gateway integration:** increase access to arXiv content.
6. **Migration to AWS:** as legacy integrations drop off, services are re-deployed in AWS.



Future: Image and Data ?

(図 13)

Submission & Moderation

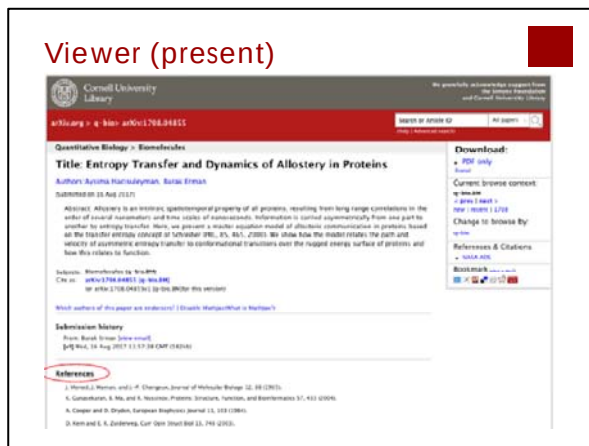


(図 15)

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。



(図 16)



(図 17)

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) というアドバイザリーボードがあって、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。

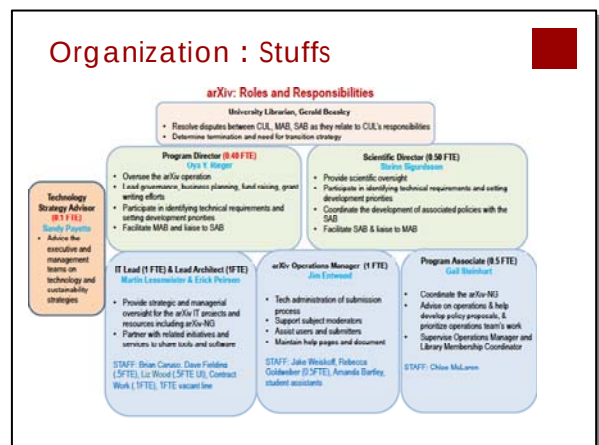
これを非常に少ない人員で行っています。図 19 の資料もオープンなものです。full-time equivalent (FTE) ですから、1 日を 1 とすれば、0.1~1 のエフォートで、エフォート管理をしながら全体を運営しています。ライブラリとこちらの仕事を共有しながらやっているというのが現状です。

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 18)



(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。

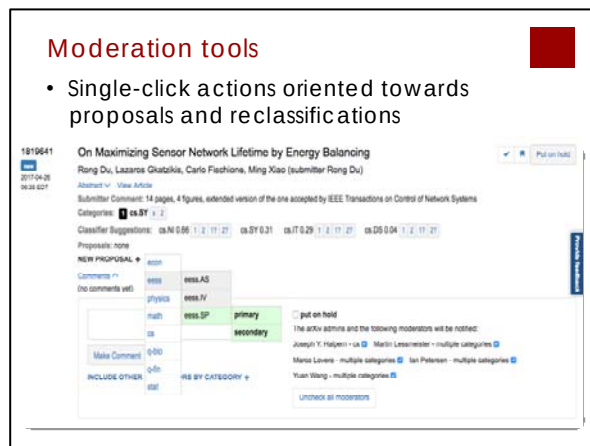
図 25 はシステムのハイライトです。いろいろなソフトを使っています。ネクストジェネレーションのアーキテクチャは、まとめると図 26 の 5 項目になります。インテグレーションが非常に重要なのですが、その前のテクノロジーを現代風書き換えるということをやっているの、それによってセキュリティの問題が当然起きてきます。セキュリティの問題に関しては今もやりとりをしていて、チェックしているという状況です。ですから、完全に確保できたという状況ではないと思っています。

図 27 は、全体の運営としてどういうサステナビリティ

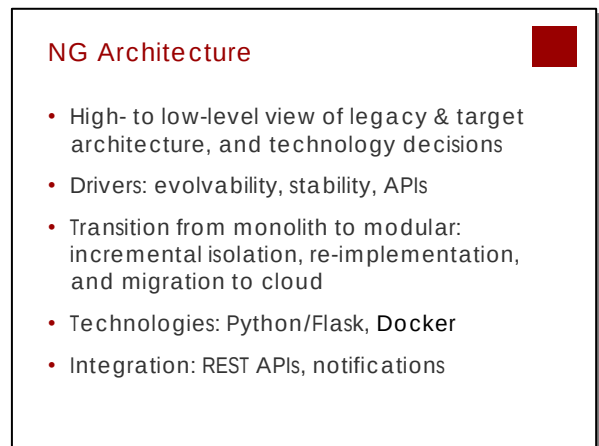
ティをどういう考え方でやればいいのかというマップです。これは Oya Rieger さんが全体の運営の中でまとめたものです。バランス良く全体を運営したいと主張しています。そのソースになる、ファンドによるお金、各コンソーシアムが支払うお金、大学が支払うお金をどう当てはめていかないといけないかが描かれています。本当は予算の関係をお見せしようと思ったのですが、ビデオで流れるので今日は割愛させていただきました。ウェブで見られますので自分で探してみてください。

結論

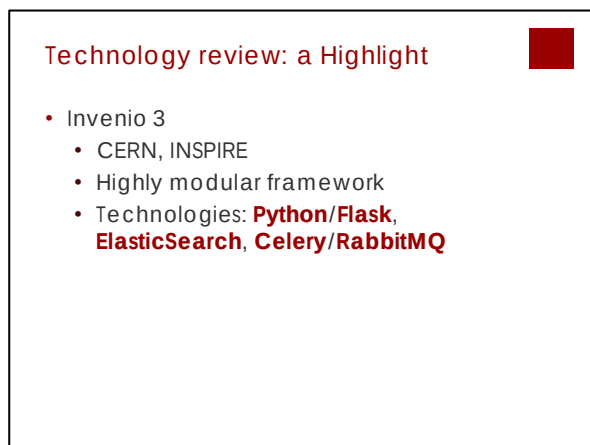
オープンライセンスに対するライセンスは何かという議論がずっと続いています。また、エンバグした後、ハイブリッドで出しているものと、もとのパブリッシュされたものとの関係がまだ微妙なところが



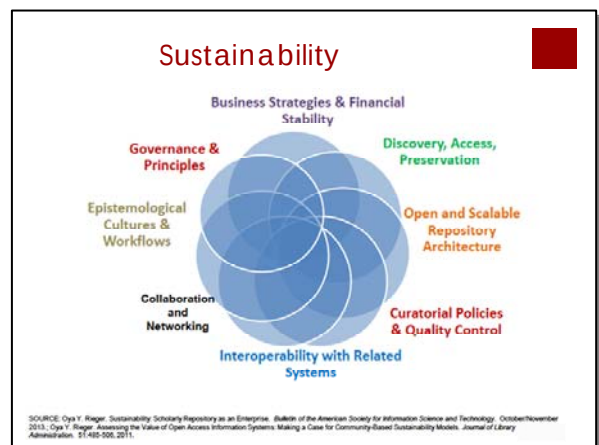
(図 24)



(図 26)



(図 25)



(図 27)

あります。エンバーゴしているからいいのではないかと思います。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があります。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。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ています。いただければよろしいかと思います。

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と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、資金はファウンダーシヨンの資金、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 年も分野を広げるということですが、コストと、分野を広げるという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ています。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います。

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だくということです。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います。

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ーの中で違う分野が起きてきているからです。それを外出ししたと厚く思っていた方がいいかと思ひます。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ー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したということです。外出ししたと同時に、もう少しそれに関連する分野を増やしていつています。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、地球物理はプレプリントサーバはなじみがなくて、私もそのようなのです。

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るのでしょうか。

●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、SCOAP³ 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。

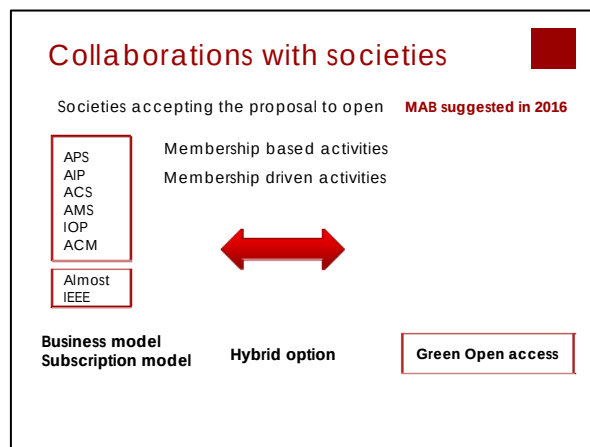
一方、先取権ということを非常にセンシティブに捉えている分野と、2〜3 年もかかるのだったら早くコミュニティに帰する方がいいと考えている分野と、ま

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で、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ください。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。