第 7 回 SPARC Japan セミナー2012

「図書館によるオープンアクセス財政支援」

arXiv.org：
オープンアクセスにおける
組織的なビジネス展望

Oya Y. Rieger
（コーネル大学 アソシエイト・ライブラリアン）

講演要旨

1991 年に設立された arXiv.org は、効果的なオンライン情報流通システムとして象徴的な地位を確立しました。また学術コミュニケーションの変遷において、リポジトリの可能性を示す良い例としてしばしば引き合いに出されます。その理由の一つとして、arXiv.org により科学的な発見を素早く世界中に普及させることが可能になり、科学をより身近なものとすることができたことが挙げられます。

2010 年、コーネル大学では arXiv.org の財政基盤を拡大する新しいビジネスモデルを採用しました。このコミュニティレベルの暫定的なモデルは多くの支援を獲得しましたが、arXiv.org のように世界中にユーザーがいる、オープンアクセスのオンラインリソースの持続可能性について、多くの疑問をも生み出しました。持続可能性とは安定したサービスを継続的に提供することです。オープンアクセスのシステムにとって、持続可能性は長期の維持管理の責任を伴います。つまり技術的、社会経済的政策、またビジネス的側面を持ち、受託責任をも含みます。オープンアクセスではその責務、管理体制、戦略的ビジネスプランを明確にし、そのようなサービスの長期的、安定的な供給を行っていかなければなりません。

Oya Y. Rieger

コーネル大学のアソシエイト・ライブラリアンで、主にdigital scholarshipや保存などを担当。図書館のデジタル化、リポジトリの発展、電子保存や電子出版、e-scholarshipなどに着目し、ニーズの評価、必要性の分析、ビジネスモデルや情報管理ポリシーの策定などに取り組んでいる。さまざまなオープンアクセス活動においても主導的な役割を果たしており、学術コミュニケーションの検証や発展に寄与している。arXiv.orgのプログラムディレクターの立場から、物理、数学、コンピュータサイエンス等のオープンアクセスの先駆けとして世界的に知られているarXiv.orgの動向を見守り続けてきた。また、コーネル大学図書館とデューク大学出版局が共同で運営しているProject Euclidの活動にも参加しており、インパクトの高い査読済の数学論文や統計データへの容易なアクセスを可能とした。METUで経済学の学位を、オクラホマ大学で行政学、コロンビア大学で情報システムの学位をとり、コーネル大学ではHuman-Computer InteractionでPh.D.を取得。

はじめに

この 20 年の間に、オープンアクセスのデジタルリポジトリは、学術コミュニケーションを支えるインフラの中でも重要な要素となってきました。オープンアクセス運動の目標は、科学知識を公共財として広めることです。デジタル出版やリポジトリを介して、学術研究を自由かつオープンに入手できるようにすることを目指しているのです。オープンアクセスは学術コミュニケーションのモデルであり、無料で制作するのとは違います。お金がかかります。オープンアクセス出

版には、スタッフや技術、時には設備を含むリソースが必要です。従ってオープンアクセスは、そのシステムの維持と発展に必要とされる健全なビジネスの原則に従わねばなりません。

arXiv の概要

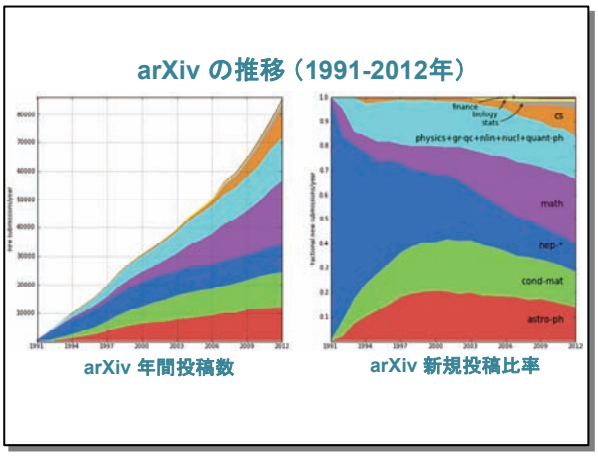
コーネル大学図書館で arXiv を運営した経験に基づき、オープンアクセスの組織的、ビジネス的なさまざまな側面についてお話しいたします。多くの機関が、研究活動や教育活動を支援するため arXiv を利用しています。arXiv は、物理学、数学、コンピュータ科学、定量生物学などを含む幾つかの分野で、科学的知見の速やかな世界的普及を促します。金銭的な障害なく自由に論文の投稿やダウンロードを行えるため、arXiv は科学をより自由なものにする上でオープンアクセスの役割を示す好例と言えます。

arXiv は、1991 年にポール・ギンスパーグによって設立され、2001 年以降はコーネル大学図書館が運営してきました。2013 年 1 月時点で、815,000 本以上の論文が arXiv に掲載されています。2012 年には 85,000 件近くの新規投稿を受け付け、世界中で 5000 万回近くダウンロードされました。2011 年の投稿数は合計 77,000 本に迫り、2011 年から 2012 年の 1 年間で、投稿数が 10%以上増えた計算になります。

1991 年の設立以来、arXiv は成長を続けています。グラフ左は、投稿論文を分野別に色分けしたもので、2012 年までの推移を示しています（図 1）。このグラフを見ると、arXiv の投稿数がどれほど増加を続けているか分かるでしょう。二つ目のグラフは、投稿論文の内訳を比率で示したものです。1991 年には投稿論文のほぼ 100%が、高エネルギー物理学だったことが分かります。2012 年には高エネルギー物理学が安定した一方、数学が大幅に伸びており、他に幾つかの分野も増えています。2012 年の投稿論文の比率を見ると、数学が 26.7%を占め、2 番目に多いテーマは固体物理学の 14.4%、次いで天体物理学の 14.3%でした。

arXiv の開発・保守に関わる予想経費として、2013～2017 年のビジネスモデルを作成しました（図 2）。2013 年について言えば、支出は 3 種類に分けられます。スタッフという項目にはプログラマーやユーザーサポートの費用が含まれ、経費は約 537,000 ドルと見積もっています。また、サーバーやネットワークなどの通信経費や他の直接費は約 74,000 ドルと見込まれます。設備のメンテナンスや管理費用、訴訟費用などの間接費も考慮に入れています。

私の上司はコーネル大学図書館員で、私は勤務時間の 15～20%を arXiv に費やしています。しかし、ここ数年は企画業務に携わっているため、arXiv に充てる時間が大幅に増えました。arXiv の管理者は 3 人いて、情報技術を担当するシメオン・ワーナーは、勤務時間の 30%を arXiv 関連の仕事に費やしています。ジェイロン・ポルチエロは、arXiv の会員プログラムの管理に勤務時間の 10%を充てています。デビッド・ラディはユーザーサポートを統括しており、勤務



（図 1）

予想経費（2013-2017年）

		内容	2013	2014	2015	2016	2017
経費							
	スタッフ						
		ユーザーサポート（スタッフ）	228,405	236,400	244,680	295,453	305,795
		ユーザーサポート（学生）	8,500	8,500	8,500	8,500	8,800
		プログラミング &システム保守	238,669	164,685	170,453	176,415	182,590
		管理& 運営	61,835	64,000	66,240	68,558	70,955
		スタッフ費用小計	537,409	473,585	489,873	548,925	568,140
	他の直接費						
		サーバーコスト	41,700	43,160	44,670	46,235	47,855
		通信経費	1,550	1,605	1,660	1,720	1,780
		スタッフのコンピュータ、ソフトウェア、消耗品	1,865	1,930	2,000	2,070	2,140
		スタッフの出張	8,280	8,570	8,870	9,180	9,500
		諮問委員会の出張補助	20,700	21,425	22,175	22,950	23,755
		直接費小計	74,095	76,690	79,375	82,155	85,030
	間接費						
		設備	67,265	60,530	62,617	69,419	71,849
		部門管理 & スタッフ支援	158,991	143,072	148,004	164,081	169,824
		間接費小計	226,256	203,602	210,622	233,500	241,673
		経費合計	**837,760**	**753,877**	**779,869**	**864,580**	**894,843**

（図 2）

時間の 20%を使っています。お気づきのように、3 人とも図書館内の別の案件にも関わっているため、勤務時間の一部のみ arXiv に携わっています。arXiv チームは、コーネル大学図書館の総合図書館管理業務の中に組み込まれているのです。例えば、デビッド・ラディは図書館の機関リポジトリや記録保管リポジトリ、数学系電子ジャーナル Euclid のユーザーサポートやポリシーなども統括しています。

持続可能性に向けた取り組み

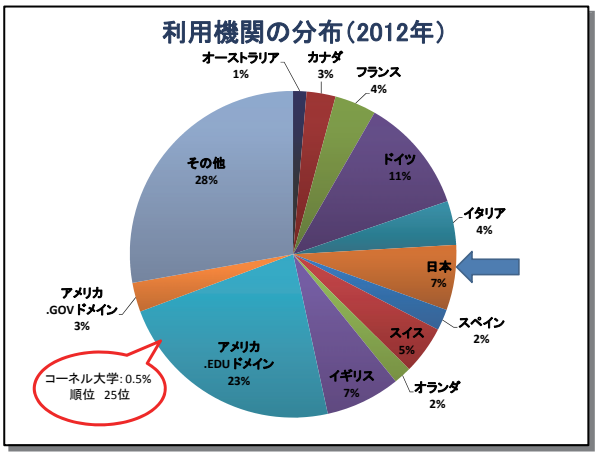
arXiv の持続可能性に向けた取り組みに関しては、arXiv の安定的な運用と継続的な発展が世界中の研究者に影響を与えます。持続可能性を確保するための戦略として、単一の機関への依存を減らそうとしています。コミュニティに支えられた幅広い基盤を持つリソースを作り出すことが目標です。2010 年 1 月以来、コーネル大学図書館は arXiv の持続可能性モデルの確立に取り組んできました。安達先生が企画段階から積極的に参加してくださったことを、ありがたく思っています。

私たちは 3 年間企画を練った末、最大の機関ユーザーである図書館と研究機関に arXiv のサポートとガバナンスに協力してもらう会員制モデルを考案しました。それぞれの会員機関に対し、arXiv 支援のため、最初の 5 年間は資金援助の約束をしていただけるようお願いするのです。これは法的な拘束力のある義務ではなく、arXiv を支援するという誠意に基づくステートメントに過ぎません。

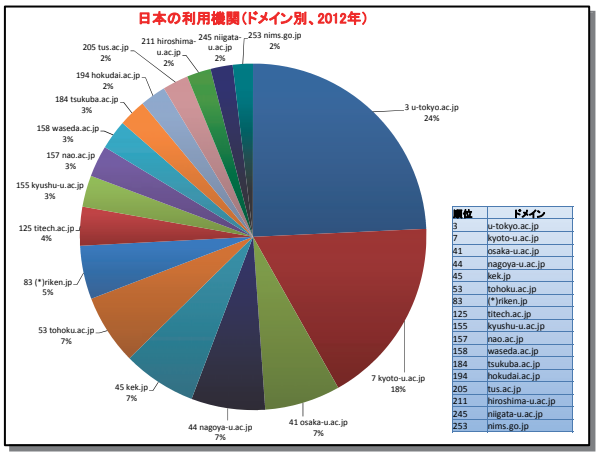
2013～2017 年のビジネスモデルには、四つの収入源があります。一つ目はコーネル大学図書館で、75,000 ドルの年間援助に加え、間接費全てについて現物出資の提供を受けており、運営コスト総額の 37%を賄っています。サイモンズ財団から 50,000 ドルの寄付も受けています。また会員機関から、1,500～3,000 ドルの範囲内で年間会費を得ています。会員機関から集めた資金の額に応じて、サイモンズ財団から年間 30 万ドルのマッチング・グラントを受けています。

arXiv のガバナンスモデルでは、ヘビーユーザーである上位 200 の学術機関・研究機関を対象としています。私たちは、会員制モデルを用いて大学や研究機関と連携することを決めました。なぜなら私たちは arXiv を運営することにより学術情報を管理し、学習や教育、研究活動を支援しているからです。arXiv のユーザーは世界中におり（図 3）、2012 年の機関別の利用分布を見ると、多くがアメリカの教育機関のドメインに由来し、これが 23%を占めます。コーネル大学の利用頻度は、世界で 25 番目に位置します。他に利用頻度が高い国として、イギリスの 7%、スイスの 5%、ドイツの 11%、日本の 7%などが挙げられます。

日本には利用頻度の高い機関が 17 あり、17 番目の機関は世界ランキングで言うと 253 番目に位置します（図 4）。世界の利用頻度ランキングで見ると、東京大学は第 3 位、京都大学は第 7 位です。日本から



（図 3）



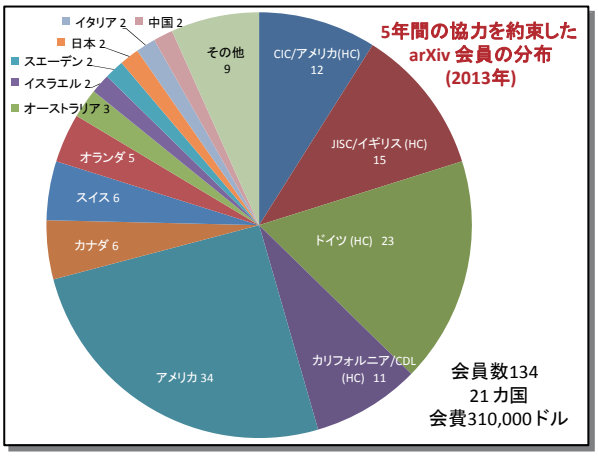
（図 4）

の利用の 24%を東京大学が占め、18%を京都大学が占めています。2013～2017 年の期間に関しては、今までに 21 カ国・134 機関が 5 年間会員になることを約束し、会費にすると合計 31 万ドルに達します（図 5）。

持続可能性に向けた取り組みで何より大切なのは、arXiv の目標と経営体制を明確に定めることでした。ガバナンス体制に関しては、コーネル大学図書館が arXiv の運用・開発に対し、管理上・財務上の全般的責任を負っています。戦略上・運用上の指針は会員諮問委員会が決めます。科学諮問委員会は、知的面の管理について arXiv に対し助言や指針を行います。実際には、会員諮問委員会は会員である機関、図書館、研究機関から構成されます。私たちはつい先日、選考プロセスを経て新たな会員諮問委員会を結成したばかりです。詳しい情報は、arXiv の持続可能性に関するウェブサイトに掲載しています。

持続可能性のための原則

コーネル大学図書館で 3 年にわたり、arXiv の未来のため持続可能なモデルの開発に取り組んだ経験をもとに、持続可能性のための原則を五つ提案したいと思います。この原則を一般化して、他のオープンアクセス事業に役立てることができれば幸いです。一つ目は、学術コミュニティと研究プロセスとの統合の深化です。各分野の特徴や研究者の研究動向は、研究者が情報システムを利用する上で重要な役割を果たします。物理



（図 5）

学者、数学者、コンピュータ科学者、生物学者はみな異なるコミュニケーションのパターン、研究パターンを持っています。arXiv は学術コミュニティから生まれたもので、技術者や図書館員が設計したものではありません。

arXiv は、学術論文の迅速な普及とオープンアクセスを提供することで、従来の出版システムを補完します。テクノロジーは重要ですが、社会的課題や文化的課題を考慮に入れる必要があります。技術のイノベーションが全てではなく、適正な技術を活用して、実際の研究動向・活動や要求に基づき研究者のニーズを支援することが大切です。arXiv が国際的に広く受け入れられている事実は、IT の改良や機能強化を行う上でユーザーの立場に立った方策をとりいれることの重要性を裏づけるものです。

二つ目の原則は、権能とガバナンス体制を明確に定義することです。arXiv のような主題別リポジトリには、長期的な受託責任を反映した明確な権能と、それに伴うガバナンス体制が必要です。ガバナンスの一般的な目的は、組織が自らの将来計画を遂行するための手段を確保することです。ガバナンスのためには、計画の持続的な実行を保証するための管理体制も求められます。優れたガバナンスとは、参加型で、コンセンサスを重視し、説明責任を負うことができ、透明性が高く、対応が早く、効率的で公平な包括的なものです。他方で、膠着状態に陥ったり、コンセンサス形成を過度に重視したりすることがないよう、鋭敏さや柔軟性も求められます。

三つ目の原則は、技術プラットフォームの安定性とイノベーションに関するものです。既存のリポジトリには、機関リポジトリ、主題別リポジトリ、記録保管リポジトリなど幾つもの種類があります。現在、状況はさらに混迷の度合いを増しつつあり、学術的なソーシャルネットワークや、研究データに重点を置いたオープンデータ運動も登場しています。持続可能な計画には、こうした多彩な背景を考慮し、arXiv のようなサービスをどうすれば全体的な枠組みの中に収められ

るかを理解しなければいけません。

特に、次の 4 点を考慮に入れる必要があります。1 点目は、他のリポジトリや関連するシステム、サービス、コミュニティとの相互運用性です。2 点目は、アクセスを容易にして見つけやすくし、ユーザーが簡単に情報を検索できるようにすることです。3 点目は、基礎データ、マルチメディアコンテンツ、研究手法などの補足的な情報をサポートする機能を盛り込むことです。4 点目は、電子媒体の長期的な保存を支援する機能を保つことです。

四つ目の原則は、体系的な開発やコンテンツの方針です。明確に定められたデータ収集方針と投稿指針が、オープンアクセス資源にプラスに働きます。質の管理も重要で、質の管理水準はデータ収集方針や投稿方針に左右されます。例えば arXiv はピアレビューを行っていませんが、論文の科学的妥当性を確認するためにモデレーターが主題別に投稿論文を評価します。オープンアクセス・リポジトリの課題の一つは、電子出版の典拠の明示と信頼性を確保することです。論文が、初回投稿からプレプリントを経て、正式な学術誌に最終的に発表されるまでの過程を、どのように追跡し整理するのか。研究データや他の補足資料などの基礎資料へのアクセスも、次第に重視されつつあります。

最後に五つ目の原則は、ビジネスプランの活用です。ビジネスプランは、営利目的の民間企業だけのものではありません。オープンアクセスにも、ビジネス戦略が必要です。ビジネスプラン作成の最大の目的は、潜在的なユーザーによるサービスまたは製品への投資を促すような、明確な価値のある提案を行うことです。価値のある提案とは、その製品やサービスによるメリットを述べたものです。言い換えると、価値提案とは、ある製品やサービスをなぜ購入し使用する必要があるかという問いに機関が答えるものです。この主眼は顧客にあるため、ステークホルダーの視点で述べられるべきです。

ビジネスモデルを通じて、財務計画も明確に伝えます。arXiv の保守、サポート、開発にはさまざまな経費が伴います。arXiv の会員制モデルのような協働ビジネスモデルでは、収入源と支出対象を把握できるよう、必要金額と予算を明確に定め正当化することが不可欠です。私たちは、オープンアクセス・システムの開発と管理に必要な経費に対し、さらに理解を深めなければなりません。ビジネスプランは、このサービスに関心を持つさまざまなグループやステークホルダーの方々に、目標や成果、課題、戦略を伝える手段でもあります。

まとめ

結論として、オープンアクセス学術コミュニケーションシステムのビジネス面を理解し、これに対処することの重要性を強調したいと思います。arXiv の事例から、多様なライフサイクルや使い勝手の問題を考慮に入れ、広い視点でリポジトリを管理する必要性が示されます。また、学術コミュニケーションをめぐる、様式や価値観の変化も考慮に入れねばなりません。オープンサイエンスやオープンデータが次第に重視されるのを受け、私たちは当初から持続可能性という要件を考慮する必要があり、今私たちが作りつつある実験途上のサービスが長期にわたる影響を持つことに留意しなければなりません。学術リポジトリをビジネスと捉えることで、新たに登場する学術コミュニケーションのシステムやその実施において、効率的、効果的で持続性のあるものになるでしょう。

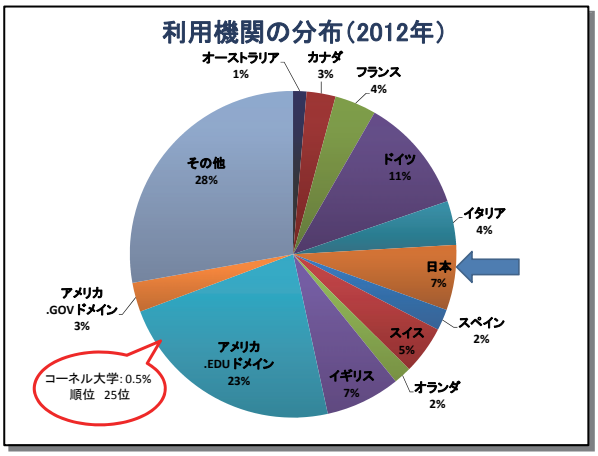
●Q1　利用機関について、国別と大学別の円グラフがありました。日本が 7%という数字からすると、サブミッションの値に近いのかと想像しますが、このユーザーは、ダウンロード利用の数字でしょうか。それともサブミッションでしょうか。もしサブミッションだとすると、これはファーストオーサーか何かでカウントされているのでしょうか。

●**リーガー**　7%は利用頻度の分布であり、機関ドメ

インでダウンロードされた論文数です（図 6）。よく投稿論文数や特定の機関から投稿された論文数を知りたいと言われますが、これは非常に難しいことです。投稿の時点では、敷居を低くするため、著者に対し最低限の情報しか求めません。少ない方がいいのです。多くの論文は著者が複数いますが、私たちは論文を提出した主要著者に対してのみ所属機関の登録を義務づけています。著者名の管理についても、標準化された方法がありません。従って、arXiv を対象に投稿ベースの統計を取るのがどれほど難しいか、想像できるでしょう。かなりの手作業が必要なので、私たちはサンプリングに基づき投稿ベースの統計を作成しています。サンプリングデータによると、利用頻度と投稿数のランキングはおおむね似通ったものになっています。

●Q1　ありがとうございました。大変クリアに分かりました。一般的な分野の投稿率と非常に近い値だったので、興味深いと思って質問しました。

●Q2　現在のコンテンツについて、種類別の比率を教えてください。Pre-print、Post-print、Published version などです。Published version は、それこそ最近ではオープンアクセスジャーナルもあるでしょうし、オーサーズチョイスのオープンアクセスもあるかと思います。およそのディストリビューションを教えてください。



（図 6）

●リーガー　arXiv に掲載された一部の論文は、複数のバージョンがあります。プレプリントが 2、3 種類、出版されたバージョン、ポストプリント、学会用バージョンもあります。一つのバージョンしかない論文もありますが、比率は分析していません。

●Q2　他の分野から自分たちの分野を新たに含んでほしいというリクエストがあると思いますが、それにどう対応されているのでしょうか。

●リーガー　他の分野から要請を受けています。今年の目標の一つは、科学諮問委員会と会員諮問委員会という二つの委員会と協議して、依頼を受けた新たな主題領域を評価するための手法を策定することです。これは非常に複雑な、時間のかかる作業です。新たなコミュニティを作るのは容易なことではありません。arXiv のスタッフは IT 担当とユーザーサポートに分けられますが、両者に同じ人数を配置しています。つまり、ユーザーサポートは私たちにとって極めて重要な課題なのです。新たなコミュニティを作る際は、訓練を受けたモデレーター、コーディネーター、スタッフを確保する必要があります。フィルタリングの手助けとして自動構文解析システムを実験的に使用していますが、これは手作業のごく一部に過ぎません。

●Q2　OA ジャーナルが増えてくる中で、分野が違うからすみ分けができるのだろうという気もしますが、どのようにお考えですか。

●リーガー　arXiv に関しては、さほど野心を持っていません。今のところ、自分たちが得意な分野に集中したいと思っています。可能性を否定はしませんが、あまり欲を出さずやるべきことに注力するよう心がけています。

●Q3　arXiv では、一般の人のアクセスはどのくらいあるのでしょうか。また、もしあるとすれば、それを

どのようにお考えでしょうか。

●**リーガー**　ユーザーに関する詳しい記録は把握していません。IP アドレスは把握していますが、ユーザーを特定するのは時として非常に困難です。特に最近では、自宅のコンピュータから arXiv にアクセスすることも増えているからです。

●Q4　図書館業務の傍らに arXiv の仕事をされているとお聞きしました。arXiv の業務に携わることで何か図書業務にメリットがあったのか、また、図書業務と兼務するのは大変だと思いますが、何かそこで工夫があるか、教えていただけますでしょうか。

●**リーガー**　大規模な業務を統括するメリットの一つは、異なる情報資源の間に幾つか共通点があることです。兼務するというのは、スタッフを配置する上で効率的な方法だと思います。例えば、私は arXiv のプログラムディレクターを務める傍ら、デビッド・ラディが率いる Euclid プロジェクトや、保存管理システム、機関リポジトリなどの関連する取り組みにも携わっています。どれも視点が非常に似通っており、別のリポジトリにもビジネスの視点を取り入れることでメリットを実感しています。私たちは持続可能なモデルの開発に取り組む中で、一つの選択肢として、内国歳入法 501（c）（3）を満たす民間非営利団体を設立することも考えました。

もしそうなっていたら、arXiv を切り離していたでしょう。arXiv は、コーネル大学とつながりを持ちつつ、大学外の組織になっていたでしょう。そうなれば、フルタイムのプログラムディレクター、IT マネージャ、ユーザーサポートが必要になったでしょう。けれどコーネル大学図書館では、既に知識を身につけ、複雑な業務を統括するための幅広い視野を備えた既存のスタッフを十分に活用できています。そのため、arXiv をコーネル大学図書館の一部として維持することに決めたのです。その方が効率的で効果的と考えたからです。

●Q5　サイモンズ財団から非常に多く資金援助を得ているというお話がありましたが、この支援は今後も継続されるのでしょうか。何か見通しがありましたら、教えてください。

●**リーガー**　サイモンズ財団はアメリカの学術振興財団で、現在 5 年間の資金援助の約束を受けています。また、最初の 5 年間の成功を受け、さらに 5 年間援助を続けるとの提案を既に受けています。サイモンズ財団は、図書館や研究機関を参加させる活動を続けることを条件に挙げています。それは、収入源を多様化する必要があると考えているからです。また、情報の管理者として図書館が責任を負うことも非常に重要だと考えています。

●Q6　arXiv におけるバージョンコントロールは誰が行っているのでしょうか。人なのか、ソフトウエアなのか、それともコントロールなどはしていなくて、ポリシーだけがあるということだったのでしょうか。聞き漏らしてしまったようなので、教えてください。

●**リーガー**　現時点でポリシーはありません。ただ、私たちが出版社や学術コミュニティ、学会と協力している以上、この課題に対処することが非常に大切だと考えています。私たちは昨年、出版社 7 社と、arXiv のコンテンツの多くを有する学会の皆さんと会議を開きましたが、彼らもバージョン管理の体系的な手法の開発に向け、私たちと協力することに大いに関心を示しました。バージョン管理はユーザーや出版社の利益につながり、この問題は、お互いに関心のある領域だと思います。企画会議以来取り立てて作業はしていませんが、私たちの計画に入っています。

●Q7　創立者のギンスパーグさんは、この arXiv にはもう関わっていらっしゃらないのでしょうか。

●**リーガー**　今も関わっていますが、数年前に彼がarXiv から手を引きたいという意向を表明したため、それに向けて取り組んでいます。彼は今、主に物理学の分野に関わっていますが、本人の意向通りポールを解放できるよう、現在、新たなスタッフ配置を試しているところです。

●**坂井**　arXiv のポリシーにクオリティコントロールというのがあったのが、私には半分意外でした。arXiv は Pre-print が出発点なので、査読付きの雑誌とは基本的に違います。ですから、科学的価値は判断しないというのが基本的な立場だと思います。恐らくおっしゃっているのは、二重投稿や盗作がないように警戒するなど、ある程度自動的にできるようなレベルのことではないかと思いますが、いかがでしょうか。

●**リーガー**　ピアレビューと質の管理は、互いに関連しているものの異なる概念です。 arXiv は、ピアレビューを行っていません。ポール・ギンスパーグによって設立されて以来、arXiv を他のリポジトリと区別する特徴の一つが、質の管理手法でした。arXiv には、130 人以上のモデレーターがいます。arXiv ウェブサイトの「ヘルプ」の中で、承認制度について説明がされていて、研究者が学術機関に所属していることを私たちが確認できるようになっています。フィルタリングも行っていて、簡単に見るだけとはいえ、論文 1 本 1 本について抄録を評価して、テーマの妥当性や、科学的な原則に則った書き方がされているかを確認します。これはピアレビューではありません。論文が一定の質を維持できるよう、修正を加えるだけです。arXiv に投稿される論文は、該当分野にとって興味深く有用な価値ある内容のものであることを期待されています。それが、中核となる原則です。