

原著論文

学術情報メディアとしての arXiv の位置づけ

The Roles and Place of arXiv in Scholarly Communication

三 根 慎 二

Shinji MINE

Résumé

Purpose: This paper identifies the roles and place of a preprint server, arXiv, in scholarly communication by analyzing all papers submitted to arXiv up to 2007.

Methods: In order to clarify 1) the growth in the annual and total number of submitted papers, 2) the percentage of papers in arXiv subsequently published in scholarly journals and their publication information, and 3) differences of years between submission to arXiv and publication in journals (i.e., immediacy), all 449,071 papers submitted to arXiv between 1991 and 2007 were analyzed by examining the following data for each paper: 1) ID, 2) “journal-ref” which is bibliographic references of publications in which the submitted paper appeared, and 3) subject categories.

Results: The main findings are as follows. 1) For 17 years, the overall annual and total number of submitted papers continued to increase, especially in astrophysics, condensed matter, and mathematics. However, the rate of increase differs between disciplines. 2) The percentage of papers in arXiv subsequently published in scholarly journals was 47.1% on average, indicating that about half of them were published. The journal-ref addition rate was much higher in physics and theoretical subject categories than in mathematics and computer science. Papers published in core journals in physics and astronomy, such as the *Physical Review* series and *Astronomical Journal*, accounted for almost half of all papers submitted to arXiv. 3) Whereas most papers submitted in high energy physics (HEP) categories can be accessed before publication, some subject categories, in contrast to HEP, generally provide access to older scholarly articles. In conclusion, arXiv is now simultaneously serving two roles: as a preprint server of journal articles and as a digital archive of scholarly articles in general.

三根慎二: 名古屋大学附属図書館研究開発室, 464-8601 愛知県名古屋市千種区不老町
Shinji MINE: Library Studies, Nagoya University, Furo-cho, Chikusa-ku, Nagoya-shi, Aichi 464-8601, Japan
e-mail: smine@nagoya-u.jp
受付日: 2008 年 10 月 21 日 改訂稿受付日: 2009 年 1 月 6 日 受理日: 2009 年 2 月 17 日

- I. 学術情報流通の電子化と arXiv
 - A. arXiv の学術情報メディアとしての属性
 - B. arXiv をめぐる言説・研究の問題点と本研究の目的
- II. 調査方法
 - A. 調査対象
 - B. 調査項目
- III. 登録論文・記事から見た arXiv の特徴
 - A. 登録論文・記事数の全体的傾向
 - B. journal-ref から見た arXiv の特徴
 - C. 登録年と掲載年の差（即時性）
- IV. 学術情報メディアとしての arXiv の位置づけ

I. 学術情報流通の電子化と arXiv

WWW の登場により、1990 年代以降から現在に至るまで、学術情報流通の電子化をめぐる議論が続いている。それは、電子化により、冊子体中心の時代において当然と見なされていた制度が見直されたり、これまで考慮されなかった新しい問題が生まれたからだと考えられる。現在、このような変革期にある学術情報流通に関する議論の中心にあるのは、冊子体の学術雑誌を電子化した電子ジャーナルである。しかし、それ以前に一種の衝撃を持って人々に認識された電子的な学術情報

メディアは、物理学分野や数学分野などのプレプリントを電子的に登録し迅速に流通させるプレプリントサーバ arXiv（アーカイブ）¹⁾ であろう。arXiv に学術情報流通の要である学術雑誌に取って代わる可能性が見いだされたこと、一部の分野とはいえ出版者や図書館を介さず、研究者間で直接論文の電子的な流通が可能になることが具体性を持って示されたことなどが、そうした認識の理由であると考えられる。

arXiv（当初は e-Print archive と呼ばれていたが、以後、名称を arXiv に統一する）は 1991 年に物理学者ギンスパーグ (Paul Ginsparg) によっ

第 1 表 arXiv に設置されている主題領域、分野およびその設置年月 (1991～2008)

	主題領域	(英語表記)	分野	設置年月
1	高エネルギー物理学格子論	hep-lat	物理学	1991 年 7 月
2	高エネルギー物理学理論	hep-th	物理学	1991 年 8 月
3	数学	math	数学	1992 年 1 月
4	高エネルギー物理学現象	hep-ph	物理学	1992 年 3 月
5	物性物理学	cond-mat	物理学	1992 年 4 月
6	宇宙物理学	astro-ph	物理学	1992 年 4 月
7	一般相対性理論・量子宇宙論	gr-qc	物理学	1992 年 7 月
8	原子核物理学理論	nucl-th	物理学	1992 年 10 月
9	計算機科学	cs	計算機科学	1993 年 1 月
10	物理学一般	physics	物理学	1994 年 10 月
11	高エネルギー物理学実験	hep-ex	物理学	1994 年 4 月
12	量子物理学	quant-ph	物理学	1994 年 12 月
13	原子核物理学実験	nucl-ex	物理学	1994 年 12 月
14	数理物理学	math-ph	物理学	1998 年 1 月
15	非線形科学	nlin	非線形科学	1998 年 4 月
16	数量的生物学	q-bio	数量的生物学	2003 年 9 月
17	統計学	stat	統計学	2007 年 2 月
18	計量ファイナンス	q-fin	計量ファイナンス	2008 年 12 月

て創設され現在に至るまで、学術情報流通の電子化の節目において何度か注目されてきた²⁾。たとえば、現在のような電子ジャーナルの提供が始まった頃である 1995 年には、創設者の名前を冠して、その学術情報流通への衝撃を反映して「ギンスパーグショック」と表現されるほどであった³⁾。1999 年には、現在の PubMed Central へとつながる E-biomed のモデルとして、arXiv が参考にされた^{4)~6)}。現在最も関心を集めているオープンアクセスの文脈においては、その実現手段の一つであるセルフアーカイビングの登録先の一つである分野型リポジトリの代表例と見なされている⁷⁾。

このように arXiv は学術情報流通の電子化が進行する各時点において常に言及されてきたが、それは arXiv 独自の機能や属性が注目され関心をひいたからと考えられる。arXiv を説明する際に、一般的には“ネット上で、プレプリント（学術誌に掲載される前の論文）を収集・公開するためのサーバー”⁸⁾といった表現がなされることが多い。しかし、以下で示すように、arXiv に関する言説を整理すると、様々なレベルにおいて arXiv とは何かについての意見・調査結果がすべて一致しているわけではない。arXiv が設立されてから 18 年の年月が経ち、2008 年現在では一日あたりのアクセス数が 60 万近くに達し（2008 年 12 月の任意の日）、対象領域も当初の高エネルギー物理学理論という物理学の一領域から、数学、計算機科学、非線形科学、数量的生物学、統計学、計量ファイナンス（計 7 分野 18 主題領域）へと年を経るごとに多様な分野へと拡大した（第 1 表）。さらに、技術的進展により学術情報流通の電子化が量質ともに進出したことを考慮すれば、現在でも arXiv は設立当初のまま変わらずプレプリントを収集・公開するためのサーバであり続けているのだろうか。もし変化していれば、何がいつ頃どのように変わり、その理由として何が考えられるのだろうか。

本研究は、arXiv に蓄積された 1991 年から 2007 年末までの 17 年分のデータに基づいて、その学術情報メディアとしての属性と変化の様相を

明らかにすることを目的とする。以下 I 章では、これまで arXiv が一般的にどのような学術情報メディアとして位置づけられてきたのか、その傾向や議論の対立点を明らかにするために、現在までになされた arXiv に対する言説および研究を整理することで、その学術情報メディアとしての属性をまとめる。II 章では研究方法を、III 章では arXiv の 17 年分の登録データを分析した結果を、IV 章で arXiv の学術情報メディアとしての位置づけに対して考察し結論を述べる。

A. arXiv の学術情報メディアとしての属性

第 2 表は、これまで arXiv に対してなされた代表的な言説および研究を概観し、複数の著者から言及されている arXiv の機能上の属性（主に、学術情報が流通、利用される際の媒体という観点から見た場合）についてまとめたものである。各著者が本文中で arXiv の属性などと明確な言葉を使っていない場合でも、文脈から arXiv について述べていることが明白であるものはすべて含めている。各セルにおいて、丸は著者が言及したことを意味しており、白丸（○）は、著者の意見・経験（本文中に明確な根拠を伴わないもの・読み取れないもの）、黒丸（●）は、調査から導きだされた結果など実証性を伴う意見を示している。

以下では、第 2 表に挙げられた各属性の中でも多くの著者によって言及されてきたものに限って説明する。その際に、各属性を提供制度、提供内容、利用行動の三つの視点から説明する。提供制度とは、arXiv が論文・記事を提供する方法、運営体制に関する属性を、提供内容は、登録されている論文・記事それ自体が持っている属性を、利用行動は、arXiv の利用者である研究者の利用行動を通して間接的に見えてくる属性を指している。

なお、安価や無料など経済的側面については多く言及されているものの、本稿の関心から外れるため、ここでは除外した。

学術情報メディアとしての arXiv の位置づけ

第2表 先行研究において

	年代	立場	提供制度			
			学術情報の 電子的流通	迅速性	広範性	安定性の 欠如
トープス ⁹⁾	1993	ジャーナリスト	○	○	○	
ギンスバーグ ¹⁰⁾	1994	物理学者	○	○	○	
ベスキ ¹¹⁾	1994	物理学者	○	○	○	
マッシュューズ ¹²⁾	1995	ジャーナリスト	○	○	○	
ハーナッド ¹³⁾	1995	研究者	○		○	
トープス ¹⁴⁾	1996	ジャーナリスト	○	○		○
ボールマン ¹⁵⁾	1996	出版社	○			○
オドリコ ¹⁶⁾	1997	数学者	○			
石山 ¹⁷⁾	1997	研究者	○	○	○	○
針谷 ¹⁸⁾	1997	物理学者	○		○	
ルッチ ¹⁹⁾	1998	研究者	○	○		
高島 ²⁰⁾ , 高島ら ²¹⁾	1998, 2000	研究者	○	○	○	○
倉田ら ²²⁾	2000	研究者	○	○	○	
トメイウオロら ²³⁾	2000	図書館員	○	○	○	○
オコンネル ²⁴⁾	2000	CERN	○			
ボイス ²⁵⁾	2000	学会	○			○
ピーク ²⁶⁾	2000	研究者	○	○		○
クイグレイ ²⁷⁾	2000	図書館員	○	●		
マクキーン ²⁸⁾	2000	図書館員	○	○	○	
ランガー ²⁹⁾	2000	物理学者	○	○	○	
ティル ³⁰⁾	2001	医学研究者	○			
ブラウン ³¹⁾	2001	研究者	○	○		
倉田 ³²⁾	2001	研究者	○	○	○	
カリヴォー ³³⁾	2001	図書館員	○	○	○	
ピンフィールド ³⁴⁾	2001	図書館員	○	○		
ガーナーら ³⁵⁾	2001	図書館員	○	○	○	
マニエール ³⁶⁾	2001	図書館員	○			○
ルース ³⁷⁾	2001	図書館員	○	○		○
デイ ³⁸⁾	2001	図書館員	○	○		
ヒルフ ³⁹⁾	2001	物理学者	○			
ラワル ⁴⁰⁾	2002	図書館員	○			
ミランダ ⁴¹⁾	2002	図書館員	○	○		
九後 ⁴²⁾	2002	物理学者	○	○	○	
高田 ⁴³⁾	2002	研究者	○	○		
ハーナッド ⁴⁴⁾	2003	研究者	○			
スティックス ²⁾	2003	ジャーナリスト	○	○	○	
バトラ ⁴⁵⁾	2003	ジャーナリスト	○	○		
nature neuroscience ⁴⁶⁾	2003	出版社	○	○		
オーデルら ⁴⁷⁾	2004	CERN	○	○		
シュワルツら ⁴⁸⁾	2004	研究者	○	●	○	
倉田 ⁴⁹⁾	2004	研究者	○	○	○	
ギンスバーグ ⁵⁰⁾	2004	物理学者	○	○		
メトカルフェ ⁵¹⁾	2005	研究者	○			
グンナルスドッティル ⁵²⁾	2005	研究者	○	○	○	
松林ら ⁵³⁾	2005	研究者	○			
ヘイ ⁵⁴⁾	2006	研究者	○			
ヘンネケンら ⁵⁵⁾	2006	天文学者	○	○		
メレ ⁵⁶⁾	2006	図書館員	○			
トーマス ⁵⁷⁾	2006	図書館員	○	○		
ハナースクラ ⁵⁸⁾	2007	研究者	○			
倉田 ⁵⁹⁾	2007	研究者	○	○	○	
デイビスら ⁶⁰⁾	2007	図書館員	○	○	○	
ギンスバーグ ⁶¹⁾	2007	物理学者	○	○	○	
アメリカ化学会 ⁶²⁾	(2001)	学会	○	○	○	○
アメリカ心理学会 ⁶³⁾	(2002)	学会	○			

注：アメリカ化学会は2001年から（Internet Archiveに基づく）、アメリカ心理学会は2002年から（同学会ホー

ムページに基づく), 方針を掲載していることが確認できる。

1. 提供制度に関する属性

a. 学術情報の電子的流通

arXiv の当初の名称が、e-Print archive であったことからわかるように、これまで紙で流通させていたプレプリントを、e-print として WWW 上で電子的に流通させるという点に、arXiv の大きな特徴がある²¹⁾。この電子的流通は、以下で述べる他の属性の前提となる基本的属性であり、二つの大きな技術的仕組みがある。

一つは、arXiv それ自体の仕組みではないが、T_EX によって論文が作成できるようになったことが挙げられる。T_EX とは、“本や論文などを印刷・電子化するためのフリーソフト”で、数式を含む論文作成ソフトの事実上の標準である⁶⁴⁾。T_EX 登場以前と以後のプレプリント作成・流通作業の劇的変化の様子は、京都大学基礎物理学研究所の九後に詳しい⁴²⁾。九後によれば、T_EX 以前はプレプリントの作成は“ひと仕事”で“手間や費用がかかった”が、T_EX の普及が“情報交換に本質的な変化”をもたらすことになる。つまり、手書き部分のないプレプリントを容量の小さなデジタルデータとして arXiv に保存し、瞬時に取り寄せることが可能になった。

二つ目は、電子的なプレプリントを、学協会・出版社・図書館を経ずに、E-mail, FTP, WWW を通して研究者個人レベルで登録・入手できるようになったことである。物理学分野の学術雑誌は天文学分野の Astrophysics Data System と並んで早期から、現在のような形態で電子ジャーナルの全文提供を行っていた先進的な分野である⁶⁵⁾。しかし、それ以前に、ギンスバーグ個人が創設した arXiv は、研究者間のみでプレプリントを電子的に流通させ、実際に多くの物理学者に利用されるようになった。それゆえ、“学術論文誌のあり方を大きく変えてしまった”¹⁷⁾と見なす者もある。arXiv によるプレプリントの流通を自らの出版物を脅かすものとして懐疑的だったアメリカ物理学会 (American Physical Society: APS) 会長リクター (Burton Richter) (当時) でさえも、2020 年までに、e-Print archive と同様に、APS の学術雑誌の目次、抄録、全文がコンピュータの端末か

ら入手できるようになるだろうと述べており⁹⁾、ギンスバーグという一研究者が考案した論文の電子的流通の仕組みを将来の学会刊行物の流通モデルとして評価していると考えられる。実際に arXiv の普及をみて、APS と英国物理学会の編集長が、ギンスバーグのオフィスに訪問をしているほどである⁶⁶⁾。

arXiv は学術情報の登録から利用の一連のプロセスをすべて電子化した初めてのメディアで、それを個人レベルで成し遂げてしまったわけであり、新しい学術情報流通モデルの可能性を具体的に示してみせた点に大きな意味があると言える²¹⁾。

b. 迅速性

ほとんどの著者によって arXiv の属性として挙げられているのが、迅速性である。arXiv に登録することによって、学術雑誌などに掲載され刊行されるよりもかなり前に研究成果を迅速に研究者集団に流通させることができる。ここでの迅速性とは、1) 刊行を待たずに、2) 一瞬にして、著者が流通させられるという点においてである。

arXiv のメインサーバを設置していたロスアラモス国立研究所 (LANL) 図書館のルース (Richard E. Luce) は、arXiv の特徴として、研究成果を個別論文という形で同僚と迅速に共有することを挙げている³⁷⁾。スタンフォード大学線形加速器センター (SLAC) のペスキン (Michael E. Peskin) は、“プレプリントが新しい結果の主要な伝達手段であり、学術雑誌に論文として掲載されることは新しい情報というよりも想起や確認として機能する”¹¹⁾と述べ、九後も、“(我々の分野では) ほぼすべての論文が e-プリント・アーカイブに投稿されるため、従来の学術雑誌に載るよりはるかに早く、世界中のあらゆる場所でリアルタイムで読むことができるようになった”⁴²⁾と述べている。

ただし、arXiv に登録された論文・記事が刊行前にどの程度早く arXiv で流通しているのかについては、以下の調査がある程度である。シュワルツらは、*Astrophysical Journal* (ApJ) に 1999 年および 2002 年それぞれの後半に掲載された論文

(各 795 編, 844 編) の arXiv への登録時期を調査している⁴⁸⁾。その結果, 論文の内容(理論系・観測系など)や分野によって差はあるが, 2004 年に確認した時点では, 大半の論文(1999 年, 2002 年それぞれの 68%, 61%)が受理後に arXiv に登録され, 次いで, 出版前後に登録されその後更新されるもの(18%, 27%), 出版前に登録されその後更新されないもの(14%, 11%)が多いことを示している。

c. 広範性

広範性とは, arXiv によってインターネットに接続できれば誰でもプレプリントが読めるようになったということである。arXiv 以前においては, プレプリントは, “ひどく不公平だった”⁹⁾という発言にも見られるように貴重なメディアであった。送付先リストにある限られた研究機関の研究者しか読むことができず, 誰もが簡単に入手し気軽に読めるものではなかった。しかし, arXiv により, “プレプリントへのアクセスがより広範なものになった”¹¹⁾ため, 研究者の職位, 機関, 国にかかわらず, インターネットを通して, プレプリントを入手できるようになった。ギンスバーグは, arXiv 創設の本来の意図は “学術雑誌に取って代わるのではなく”¹⁰⁾, “刊行前の資料 (prepublication materials) に対する平等かつ均一な世界規模のアクセスを提供すること”⁵⁰⁾であると述べている。

arXiv がもたらした利用者の世界的広がりを示す具体的なデータは現時点では見当たらず, 参照できるものとしては登録者のメールアドレスの TLDs の分布が挙げられる。TLDs を基準にすると, 2005 年から 2007 年までに登録された 15 万件以上の論文・記事の分布は 135 カ国にもなる (.edu: 23.1%, .de: 8.6%, .fr: 7.4%, .it: 5.6%, .uk: 5.5%, .jp: 4.8%, .com: 3.9%)⁶⁷⁾。

d. 安定性の欠如

これまで学術情報の実質的な流通は学会, 商業出版社, 図書館がそれぞれ役割を担ってきた。しかし, arXiv は, ギンスバーグの研究室におかれたサーバが e-print の登録, 流通, 保存すべてを担っていた “私的なサービス”⁶⁸⁾にすぎなかった

(現在はコーネル大学図書館の管理下にある⁶⁹⁾)。こうした, これまでの慣例と対極に位置づけられる arXiv に対しては, その安定性のなさ, 具体的には保存体制と版管理が問題点として指摘されている。たとえば, 物理学者のボーデンシャッツ (Eberhard Bodenschatz) は学会 (APS) の立場から, “ギンスバーグのプレプリントサーバは一人の人間が運用しており, もし彼が明日すべてのマシンの電源を切ると決めたら, 全てが消え去ってしまう。彼がそうしないと誰が保証できるのか。一個人よりも学会によるほうがおそらくうまくやるだろう”¹⁴⁾と述べている。アメリカ化学会では, 複数の改訂と更新の乱用, 継続性の欠如と長期的アーカイビングの非信頼性を問題として挙げている。商業出版社アカデミックプレス (当時) のボールマンも, arXiv は半永久保存が未保証であり, 異なる版の管理も不十分であると述べている¹⁵⁾。

一方で, 先のルールは, arXiv の保存体制と版管理は問題ではないとしている³⁷⁾。著者が気まぐれ (whim) で論文・記事を登録・削除するためアーカイブとしてシステムが不安定であるということは神話であると述べている。ミラーサーバが世界中に設置されており (分散して保存されている), 版管理についても, いったん arXiv に論文が登録されると, タイムスタンプが付与され, さらに変更を加えると新しい版が作成され, 前の版に対してもアクセスできるからである。たとえ現在の版が取り下げられても, 以前の版は依然として検索可能であるとしている (登録した論文・記事を完全に取り下げることとはできず, 本文は削除されるが, 取り下げた事実は残る)⁷⁰⁾。

arXiv の安定性欠如に対する懸念は, 2000 年以前には言及されることもあったが, その後ほとんど見られなくなっている。これは, 運営体制や arXiv を背後で動かしているプログラミングの性能向上が安定性の問題を解決したからだと考えられる。具体的には, 複数のミラーサーバが世界中に設置されることで分散保存が可能なり, arXiv というシステムが世界中の研究機関の協力体制のもと運営されるようになった。さらに, 1997 年

以降、論文の各版も記録・保存できるようになり、システム全体として安定性が増していった。当初はロスアラモス国立研究所の研究室にあったサーバが、2001年にコーネル大学に移り同大学図書館から支援を受けたことで、安定した運営の継続性が見込まれるようになったことが最も大きいと考えられる。以上のような運営体制が整えられたことが、結果として先のような否定的意見が見られなくなった理由だろう。

2. 提供内容に関する属性

a. 査読の有無

arXivには学術雑誌論文だけでなく、図書や一般雑誌など他の媒体に発表された多様な論文および記事が登録されていると言われている。本稿では、これらを総称して「論文・記事」とし、学術雑誌を発表媒体とする刊行前の原稿を電子化したプレプリントに対してのみ「e-print」という言葉を用いる。

インフォーマル・コミュニケーションとフォーマル・コミュニケーションは、学術雑誌を基準として峻別されるわけだが、それは査読による品質管理があるからであり、言い換えれば、査読を通ったもの以外は流通しないということである。しかし、arXivは査読を受けていないものも査読を通ったものも区分することなく、すべて流通させているのが現状であると考えられる。

arXivに登録されたこれらの多種多様な論文・記事の玉石混淆といった状態をめぐって、学術雑誌を刊行する立場の学会および出版社側と論文・記事を利用する研究者側とで意見の食違いが見られ、以下では、双方の立場の意見を紹介する。

アメリカ天文学会のボイス (Peter Boyce) は、学術雑誌が持つ5機能との比較で、プレプリントサーバを位置づけている。ボイスは、プレプリントサーバは現状把握 (status: 研究者集団を最新の状態に保つ) と報知 (news: 最新の研究成果の流通) の二つの機能を満たすが、より重要である残りの3機能 (情報 (information: 知識の蓄積)、著者の評価 (author evaluation)、歴史 (historical: 科学の進展の記録維持)) については現時点

では満たしておらず、決して満たすこともないだろうとしている²⁵⁾。商業出版社の立場から、ボルマンは、arXivはあくまでプレプリントのデータベースでしかなく、業績の認定にならない品質保証のない論文を流通させている点を問題点として指摘し、arXivという「ウィルス」が高エネルギー物理学分野の集合的記憶を腐敗させると述べている¹⁵⁾。

プレプリントの位置づけは、学会の刊行する学術雑誌の編集委員会の方針にも現れている。たとえば、アメリカ化学会 (ACS) は、プレプリントに関する編集委員会の方針を掲載している。その中で、プレプリントサーバの欠点として、質の低い・繰り返しの出版物があふれ信頼でき価値のある情報の発見が困難になる可能性、査読の欠如などを挙げて、ACSはプレプリントサーバへの登録を電子的出版と見なし、登録済みの論文が学術雑誌に投稿されても同学会の学術雑誌には載せないこと、プレプリントの引用は査読を経ない資料の引用として見なすことを明記している⁶²⁾。アメリカ心理学会は、「インターネット上における論文の公開」というページを掲載しており、未公開論文のインターネット上の公表を禁止してはいないものの、そのような論文を投稿する際には編集者に知らせる義務があること、編集者によっては既に刊行したものと見なしたり、査読しないこともあると明記している⁶³⁾。

一方で、研究者の立場から、ギンスバーグは、arXivは成り行き任せのUseNetのニュースグループでも、誰でも無料で読めるブログのようなものでも決してないと述べている。登録される論文・記事は、少なくとも査読にかけることができる品質のものが登録され、結果としていずれかの学術雑誌の編集要件を満たしうるし、満たすであろうし、満たしており、そのための様々なスクリーニングを行っているため、読者がarXivを有益なものだと見なしていると考えている^{50), 61)}。物理学者の九後も、“最近の参考文献に関してはそれが学術雑誌に掲載されている方がいいが、全てe-プリント・アーカイブからdownloadする”⁴²⁾と述べている。実際の利用者である物理学

者にとっては、学会や出版社関係者が懸念するほど、査読を通っていない論文・記事であっても信頼性という点で問題ないとされていることが伺える。それは、1) 論文の善し悪しは研究者が判断できるという自負や、2) 高エネルギー物理学分野(特に共同研究)では原稿が外部に公表されるためには、共同研究グループの出版委員会の承認が必要であり、その時点で多数の研究者のチェックがなされること、3) 受理されるまで投稿し続けるため、領域によってはプレプリントがポストプリントに近いことがあるからだと推測される^{5), 44), 71)}。

以上から、学会・出版社側は、情報の流通媒体としては学術雑誌が絶対であり、査読を経ない論文を流通させることそれ自体が認められないという立場であると思われる。一方で、研究者は、上記のような理由で、arXivに登録された論文・記事は最終的に学術雑誌に掲載されるものに近いのであるから、入手時点での査読の有無も学術雑誌以外で論文・記事を流通させることも問題視していないことが伺える。両者の間には、arXivに登録された論文・記事の位置づけに大きな開きがあると言える。

b. 登録論文・記事の種類

このような意見の食違いをもたらすarXivに登録された論文・記事であるが、実際にarXivには学術雑誌掲載論文のプレプリントがどれだけ登録されているのか。よく引用されるものとして、SLACが運営しているSPIRESのデータベース管理者(当時)であるオコンネルの発言(2002年当時)がある²⁴⁾。彼は登録されているものの7割が学術雑誌論文で、2割が会議録論文だと述べているが、いつの時点の状況なのかさらにはその具体的根拠も示されていない。ハードは、私信として、1996年当時、高エネルギー物理学現象、高エネルギー物理学理論、高エネルギー物理学格子論、一般相対性理論-量子宇宙論の4主題領域において、arXivに登録された論文・記事のうち学術雑誌論文が占める割合はそれぞれ52.7%, 55.3%, 51.3%, 46.4%であると述べており⁷²⁾、オコンネルと比べるとかなり低い数値になっている。

る。

より実証的なデータを伴うものとして、高島²⁰⁾、高島と倉田²¹⁾、メレ(Salvatore Mele)ら⁵⁶⁾の調査がある。高島、そして高島と倉田は、1997年までにarXivに登録されたすべての論文・記事約8万件を対象に、journal-refの調査を行っている。journal-refとは、登録した論文・記事の後の掲載先を示すもので、原則として確定した論文の書誌事項が記載されている⁷³⁾。journal-refが付与されていれば、その論文・記事は学術雑誌に掲載されたと一般に考えることができる。付与率は、最高で高エネルギー物理学格子論が81.5%、最低で物性物理の20.2%となっている。

メレらは、2005年までにarXivの高エネルギー物理学4主題領域(hep-ex, hep-lat, hep-ph, hep-th)に登録されたすべての論文・記事約10万件を対象に分析を行っており、学術雑誌掲載率(journal-refの掲載先として会議録論文、講義ノート、博士論文、図書があるものは除く)を算出している。その結果、高エネルギー物理学の4分野全体でみると、1991年に75%、1998年に66%、2005年には58%となっており、年々わずかながら学術雑誌掲載率が低下していることが示されている。

以上から、arXivに登録されている論文・記事は、主題領域によって学術雑誌論文が占める割合が異なっていることが言え、必ずしもすべてが学術雑誌に掲載されているわけではないこと、掲載率についても常に一定ではないことが推測されるが、最近の動向については高エネルギー物理学分野以外は明らかではない。

c. 登録率

比較的狭い範囲の主題領域に関連する学術雑誌掲載論文が“過不足なく”³²⁾集中して登録されている点もarXivの大きな特徴であると言われている。すなわち、arXivを見れば、様々な学術雑誌に分散して掲載されていると思われる自分の研究領域に関連するほぼすべての論文が、特定の主題領域のアーカイブに登録され入手できると考えられている。arXivにおける分類は、一般的なデータベースによく見られる学術雑誌のタイトル

別や大まかな学問分野ではなく、高エネルギー物理学を例にとると、理論、実験、現象、格子論など特定領域が設定されており、関連する論文の大部分が登録されている領域もあると考えられている。

ギンスパーグは、こうした状態を指して、arXiv は学術雑誌 1 タイトルでは扱うことができないような広がりを持っていると述べている⁵⁰⁾。九後の“(我々の分野では) ほぼすべての論文が e-プリント・アーカイブに投稿される”⁴²⁾ という発言、ハーナッドの“高エネルギー物理学分野におけるすべての年次成果はそこ(著者注: arXiv)に蓄積される”⁴⁴⁾ という発言、スティックスの“高エネルギー物理学分野の出来事が漏れなく記録されている”²⁾ という発言も、同様の立場であると言える。

arXiv に登録された論文・記事が、結果としてどのような学術雑誌にどれだけ掲載されているかについては、実証的なデータを示すものとして、高島、そして高島と倉田による調査がある。arXiv は複数の主題領域において投稿先の上位にくる学術雑誌が重複しており、既存の学術雑誌よりも狭い主題領域をカバーしていることから、結果として arXiv を使えば複数の雑誌に掲載される自分の関連領域の論文を一度に入手できることを見いだしている^{20), 21)}。

逆に、学術雑誌に掲載された論文が arXiv に登録されている割合を調査した研究もいくつか見られる(第 3 表)。物理学や天文学分野の学術雑誌

においては、掲載論文が arXiv に登録される割合は年々上昇しており、タイトルによっては 9 割にも及んでいるが、同じ分野においても下位領域によって登録率はばらつきがあることがわかる。しかし、他分野や arXiv 全体については最近の傾向はわかっていない。

3. 利用行動から見た属性

研究者の arXiv の利用行動それ自体はメディアの属性ではないが、行動を通して見えてくる属性もある。ギンスパーグが指摘するように、物理学以外の統計学や数量的生物学などといった主題領域も arXiv に設置されているが、プレプリントを電子的に流通する arXiv というシステムは、それを導入すればどの分野においても誰もが利用するようなものではないとも指摘されている。倉田と松林は、日本の物理学研究者に対する質問紙調査を 1998 年および 2003 年に実施しており、研究者の利用行動や認識を通して、arXiv は物理学者の中でも若手や理論研究者など独特の利用者層を持つ情報メディアであると位置づけている^{22), 53), 59)}。

このように限定的な普及が見られる理由の一つとして、高エネルギー物理学のような領域においては、電子的にプレプリントを流通させるための土壌、すなわちプレプリント文化が arXiv 創設以前から存在していたことが大きい。数学者で電子出版に関する研究も行っているオドレイコ(Andrew M. Odlyzko)は、arXiv が高エネルギー

第 3 表 学術雑誌掲載論文の arXiv 登録率

著者名	調査年	調査対象	調査対象の刊行年	登録率
高島 ²⁰⁾ , 高島ら ²¹⁾	1997	e-print の投稿先学術雑誌	1996	3.3%~87.4% (推定)
シュワルツら ⁴⁸⁾	2004	Astrophysical Journal (ApJ) 同タイトルの下位領域別	1999, 2002 1999	61%, 72% 16.0%~88.2%
メトカルフェ ⁵¹⁾	2004	Monthly Notices of the Royal Astronomical Society, ApJ, Astronomy & Astrophysics	1992~2004	約 60%~80% (2004 年時)
ヘンネケンら ⁵⁵⁾	2004	天文学と物理学分野の学術雑誌 15 タイトル	1992~2004	約 40%~90% (2004 年時)

物理学において急速に普及し情報入手における第一の情報源となっているが、それは同分野においては 80 年代 (60 年代と考えられるが、オドルイコは 80 年代と言っている) のままにする) に大量の紙のプレプリントを流通させる文化を発展させてきたからであるし、実際に arXiv に登録されるプレプリントは従来の学術雑誌に発表されていると述べている。よって、プレプリント文化のない他の領域までにこのような情報メディアが普及するとは考えがたいとしている¹⁶⁾。

B. arXiv をめぐる言説・研究の問題点と本研究の目的

1. arXiv をめぐる言説・研究の問題点

これまで、arXiv をめぐる過去の言説と研究を、提供制度、提供内容、利用行動の 3 点から見てきたが、これらから arXiv とは何か、その学術情報メディアとしての位置づけを理解しようとする際に、問題となってくる点は、大きく分けて二つあると考える。

一つは、arXiv に対する意見に相違があり、必ずしも一様ではないことである。これらの言説・研究を見る限り、以下の 4 点について意見の相違が見られ、arXiv とは「ネット上で、プレプリント (学術誌に掲載される前の論文) を収集・公開するためのサーバー」⁸⁾ といった簡潔な表現以上の様態を持つ、より複雑な学術情報メディアであると考えられる。

第一に、e-print とは、「学術雑誌」に投稿あるいは掲載されたものかについての相違がある。すなわち、arXiv に登録されている論文・記事は、現在でも e-print がほぼ全てであると考えてよいのかという問題である。学術雑誌以外に掲載された論文や記事が含まれているとすれば、どの程度でありどのような種類のものなのか明確ではない。第二に、登録されたものすべてが最終的に学術雑誌に査読を通り「掲載されるのか」についての相違である。学会や出版社関係者の中には、あたかも arXiv に登録されているものすべてが査読を経ないものであるかのように主張してい

る者もいる一方で、ギンスバーグは登録される論文・記事は、少なくとも査読にかけることができる品質のものが登録され、結果としていずれかの学術雑誌の編集要件を満たすものだと主張している。学術雑誌への掲載を示す journal-ref の付与率を調べた先行研究についても、創設から最近にわたって全主題領域における付与率の傾向を示しきれていない。第三に、学術雑誌に掲載される前に、arXiv に登録されているのかについての相違である。シュワルツらの先行研究が示しているように登録論文・記事がいつ登録されるかは、同一分野でも下位領域によって異なっている。技術的には、掲載後の論文も登録することも可能であるが、実態は不明である。最後に、誰が利用しているのかについて意見の相違がある。つまり、プレプリント文化がなくても arXiv に主題領域が設置されているという意見と、物理学においてさえも arXiv を使う人は限られているという意見との間には、arXiv をどう位置づけるかに大きな開きがあるように見られる。以上の 4 点について意見の相違があるということは、arXiv とは何か、その学術情報メディアとしての位置づけは明確になっていないと言える。

二つ目は、こうした位置づけの不明確さの裏返しとして、全般的に議論の根拠に実証性が乏しいことが挙げられる。第 2 表で示したように、arXiv に関する論文・記事は、意見が大半を占め、議論の元となるデータが十分提供されているとは言えない。調査研究に基づいたものは、松林と倉田⁵³⁾が指摘しているように、少数にとどまっている。調査研究についても、先に示したように、対象となる年代や分野 (主に物理学や天文学) が限られており、18 年間に 18 主題領域に登録された 50 万件以上の論文・記事という arXiv の広がりをも十分にとらえきれていない。arXiv とは何か、どのような学術情報メディアであるのかを理解しようとするのであれば、このような広がりを捉えることができる調査が求められる。たとえば、先のメレらの調査は、高エネルギー物理学 4 領域に登録された論文・記事の学術雑誌掲載率が漸減していることを示しているが、これはその分

野のデータすべてを分析して初めて言えることである。よって、arXiv をめぐる議論の相違を解消し、その学術情報メディアとしての位置づけを明らかにするためには、可能な限り実証的なデータに基づいて、領域横断的かつ経年変化をとらえることができる調査が必要であると考ええる。

2. 本研究の目的と仮説

以上の arXiv の言説・調査研究が抱えている問題点に対して、本研究は、arXiv に登録されたすべての論文・記事を調査するというアプローチをとる。1991 年から 2007 年末までの期間で 17 主題領域に登録された論文・記事約 45 万件すべてを分析することによって、arXiv が持つ広がりや損なうことなく、その学術情報メディアとしての位置づけを明らかにできると考えるからである。本研究では、まず登録データの分析を通して全体像を把握することを優先し、先に示した四つの意見の相違のうち最初の 3 点に焦点を当て、残りの研究者の利用行動から arXiv の属性を見る調査は行わない。

このような問題意識とアプローチのもとで、本研究は、以下の三つの仮説を設定し、全登録データの調査分析を通して検討することで、arXiv の学術情報メディアとしての位置づけを示すことを目的とする。

1. arXiv に登録されたものは、最終的に何らかの成果公表メディアに掲載される
2. その掲載先は、主に学術雑誌である
3. arXiv に登録されたものは、プレプリントであり学術雑誌に掲載される以前からアクセス可能である。すでに学術雑誌に掲載済みの論文が登録されることはない

II. 調査方法

A. 調査対象

arXiv に登録されているすべての論文・記事に関するメタデータを、Open Archives Initiative Protocol for Metadata Harvesting (以下、OAI-PMH) の ListRecords を用いて 2008 年 1 月に

ハーベスティングした。本研究で対象となる年代範囲は、1991 年 8 月から 2007 年 12 月 31 日までに登録された全 449,071 件であり、この期間のデータが本研究の基礎データとなる。OAI-PMH を通して得られるメタデータからは、登録された各論文・記事について、ID (id), 主題領域 (setSpec), タイトル (title), 著者名 (author), コメント (comments) (任意), journal-ref (journal-ref) (任意), 登録カテゴリ (categories), 抄録 (abstract) を入手することができる。主題領域と登録カテゴリは、前者が論文・記事の主とする専門領域を、後者は前者の中に含まれる下位領域および関連するその他の主題領域 (cross-ref) を、それぞれ示している。なお、第 1 表に示したように arXiv には 2008 年末時点で 7 分野 18 の主題領域のアーカイブが設置されているが、本稿では 2007 年末までの計量ファイナンスを除いた 6 分野 17 主題領域を対象としている。

B. 調査項目

本研究では、各論文・記事に付与された 1) ID, 2) 主題領域, 3) journal-ref, 4) 登録カテゴリを分析対象とした。なお、コメントには通常、その論文・記事の図表数や総ページ数だけでなく、まれに journal-ref に相当する情報も含まれることもあるが、原則として調査対象外とし、journal-ref に欠けている情報を補える場合のみ参照した。

先に設定した三つの仮説を検討するにあたって、上述した 4 項目から得られたデータをもとに、以下の点について調査を行った。

まず、各論文・記事に対して journal-ref の有無を確認した (仮説 1)。journal-ref は「Nucl. Phys. B 368 (1992) 444-462」のように入力されており、この場合、1992 年に刊行された *Nuclear Physics B* の Vol. 368 の 444 ページから 462 ページに掲載されたことを示している。journal-ref は登録者である研究者が入力するもので品質管理されていないため信頼性に欠けるという指摘もあるが³⁶⁾、SLAC と協同で管理がなされることで⁵²⁾、大部分のものが統一された形式で掲載先タイトル、刊行年、巻号、ページ数が記載されてお

り、一定の信頼性があるものと考えられる。本稿では、調査中に気づいた明らかな間違いは修正し、それ以外は、そのまま利用している。

次に、journal-refに含まれている掲載先のタイトルごとに集計し、加えて主題領域ごとに同様の集計を行った（仮説2）。

さらに、arXivに論文・記事を登録することがどれだけ普及しているかを推定するために、各主題領域において総登録数が多い上位10タイトルを対象に、登録された論文・記事が各掲載先タイトルの総掲載論文数（一年単位）に占める割合を算出した。各タイトルの年間総掲載論文数は、INSPECを利用して算出した。対象タイトルがINSPECに収録されていない場合やデータに明らかな不備がある場合に限り、Web of Science、MathSciNet、PubMedなどのデータベースや、APS JournalsやScienceDirectのような電子ジャーナル提供元のサイトを利用して確認した。

最後に、arXivに登録された年とjournal-refに付与された書誌事項に記載されている刊行年の差をとることにより、その論文・記事が「刊行年

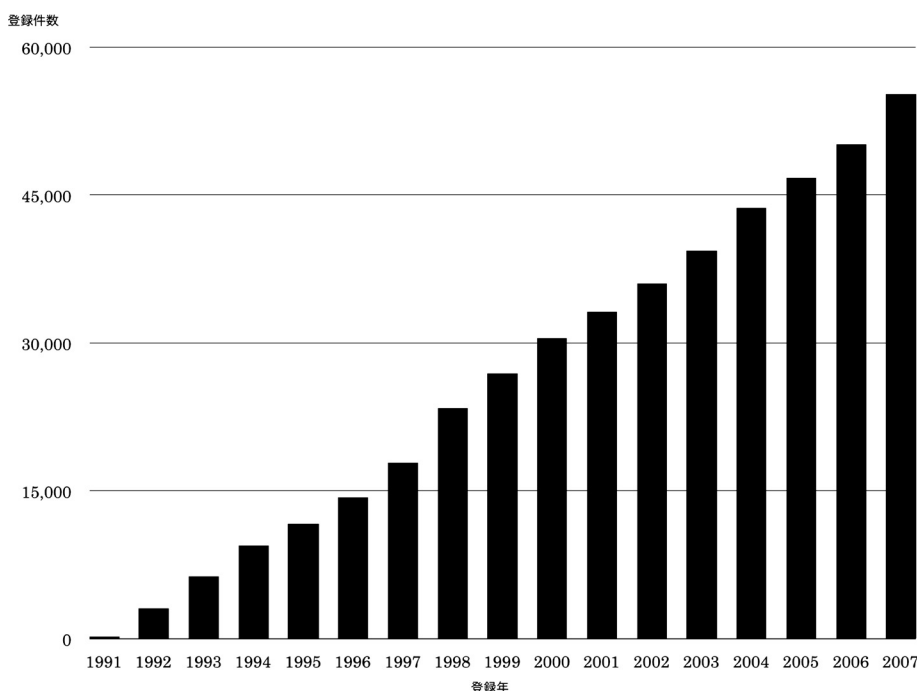
より前に登録」、「刊行年と同年に登録」、「刊行後に登録」されたかを確認した（仮説3）。論文・記事に付与されたIDには、「hep-th/9108001」といったような識別子が付与されており、これはhep-th（高エネルギー物理学理論）のアーカイブに1991年8月に登録された1番目のものを意味している。

III. 登録論文・記事から見たarXivの特徴

A. 登録論文・記事数の全体的傾向

まずarXiv全体の傾向を見ることで、どのような学術情報メディアであるかを概観する。第1図は、1991年から2007年までに登録されたすべての論文・記事について、年ごとに登録件数を示したものである。2007年には、登録件数が5万5千件を超え、10年前の3倍以上になっていることから示されるように、年間登録件数は一貫して増加傾向にあり、一度も前年度を下回ったことがない。

第4表に示すように、各主題領域に登録された論文・記事の件数を見ると、物性物理学が全体の



第1図 arXivの年度別登録件数の推移

2 割弱を占めており、宇宙物理学、数学、高エネルギー物理学現象、高エネルギー物理学理論と続いている。登録件数が多い主題領域は、arXiv が創設されてまもなく設置されているが、逆に計算機科学 (1993)、高エネルギー物理学格子論 (1991)、高エネルギー物理学実験 (1994)、原子核物理学実験 (1994) のように設置時期が早くても、登録数が少ない領域も存在しており、主題領域間における利用には依然として差異がある。このように登録数は設置年にかかわらず主題領域で差が大きい、大半の主題領域において登録数が毎年増えており (第 5 表)、特に宇宙物理学、物性物理学、数学は、年間の登録件数と増加率が他の主題領域と比べて高い。第 2 図は、登録件数上位 4 主題領域の 17 年間における登録件数の推移を示したものであるが、arXiv の中心的主題領域と言われてきた高エネルギー物理学は、4 領域 (hep-th, hep-ph, hep-ex, hep-lat のこと。第 2 図では 4 領域を合わせて高エネルギー物理学 4 領域としている) を合わせれば総登録数自体は現在でも最多であり、年間総登録件数も全主題領域の中でも最多である期間が長かった。しかし、2000 年あたりから安定傾向に入り、2005 年には 4 領域を合

わせても宇宙物理学と物性物理学に抜かれ、2006 年には数学よりも少なくなっている。年間登録数が最も多かった高エネルギー物理学現象も、1998 年の時点で宇宙物理学に抜かれている (第 5 表を参照)。宇宙物理学、数学、物性物理学といった主題領域では年間 1 万件前後登録されており、年間登録件数や増加率という点から見れば、今後、これらの領域が arXiv の中心的主題領域になる可能性がある。

B. journal-ref から見た arXiv の特徴

1. 付与率

journal-ref の付与率は、17 主題領域、17 年間全体で 47.1% である (第 6 表)。2006 年と 2007 年は、それ以前と比べて付与率が低いため (第 3 図)、2005 年までに限定すると、52.7% まで上昇する。これは、arXiv 登録時点で受理されていないものが多く含まれているためと推測される。全体的な傾向として、journal-ref の付与率は年々増加するということではなく、98 年以降ほぼ安定傾向にあり、年度をさかのぼるにつれて付与率が高くなっている。journal-ref の付与率から判断

第 4 表 各主題領域における論文・記事の登録件数と全体における割合 (1991~2007)

	主題領域	設置年	登録件数	シェア
1	物性物理学	1992	86,281	19.2%
2	宇宙物理学	1992	84,052	18.7%
3	数学	1992	58,014	12.9%
4	高エネルギー物理学現象	1992	55,364	12.3%
5	高エネルギー物理学理論	1991	45,715	10.2%
6	量子物理学	1994	23,265	5.2%
7	物理学一般	1994	18,977	4.2%
8	一般相対性理論・量子宇宙論	1992	18,313	4.1%
9	原子核物理学理論	1992	13,212	2.9%
10	計算機科学	1993	9,906	2.2%
11	高エネルギー物理学格子論	1991	8,499	1.9%
12	高エネルギー物理学実験	1994	8,362	1.9%
13	数理物理学	1998	7,023	1.6%
14	非線形科学	1998	5,897	1.3%
15	原子核物理学実験	1994	3,405	0.8%
16	数量的生物学	2003	2,521	0.6%
17	統計学	2007	265	0.1%
	合計		449,071	100.0%

すると、arXiv に登録された論文・記事の約半数強が e-print であることがわかる。ただし、登録者が学術雑誌等に掲載されても journal-ref を更新していないことが十分考えられるため、付与率（学術雑誌掲載率）を低く見積もっている可能性がある。

journal-ref の付与率も、登録数と同様に主題領域で大きな差が見られる（第 6 表）。付与率が高いのは、2005 年までのデータに限定した場合、高エネルギー物理学理論の 77.7%，高エネルギー物理学格子論の 77.1%，原子核物理学実験の 75.1%，原子核物理学理論の 74.9%，一般相対性理論・量子宇宙論の 73.0%，高エネルギー物理学現象の 70.3% など、実験系の原子核物理学実験を除き物理学の理論系が多い。逆に付与率が低いのは、下から数学の 24.2%，計算機科学の 31.2%，数量的生物学の 33.3%，非線形科学の 36.0%，物理学一般の 36.4% であり、物理学系以外の主題領域で低い傾向にあることがわかる。

journal-ref の付与率が高い主題領域では、高エネルギー物理学格子論を除いて、設置後数年は不安定な時期が見られるが、その割合が経年で大きく変化することはない。一方で、付与率がそれほど高くない主題領域では、経年変化が見られるところもある。たとえば、物性物理学では、付与率が 1994 年（21.5%）から 2005 年（56.5%）にかけて上昇しており、逆に数学では 1992 年は 41.2% であったが 2005 年には 20.0% まで減少している。宇宙物理学では、1992 年から 1997 年にかけて付与率が 71.2% から 22.8% まで減少しているが、2003 年には 67.8% まで上昇し、近年では 5 割から 6 割前後を保っている（第 5 表）。このことから、学術雑誌の原稿を arXiv に登録するという行為は、主題領域で見ると、物性物理学で普及の傾向が見られる以外は、年を経ても普及していないものと考えられる。

journal-ref の付与率から判断する限り、高エネルギー物理学分野においてさえも、arXiv に登録されたすべての論文・記事がその後学術雑誌に掲載されるということではなく、journal-ref が付与されない学術雑誌掲載論文以外の論文が登録さ

れていると考えられる。付与率が主題領域によって大きく異なることから、たとえば、高エネルギー物理学理論と数学とでは、journal-ref の意味合いをはじめとして、arXiv の位置づけがそれぞれの研究者集団内において異なっていると推測される。

付与率が 2006 年と 2007 年でその他の年よりも大きく低下しているがそれ以前はほぼ一定を保っていることから、arXiv に登録される論文・記事のうち、登録時点では受理されたものは少数で大半は受理されておらず、登録後 1 から 2 年で査読を通り受理されたものとして、journal-ref が更新されるものと考えられる。

2. 掲載先

journal-ref が付与された論文・記事が、結果としてどのような学術雑誌にどれだけ掲載されたかを調査した。具体的には、journal-ref を元に、1) arXiv 全体および 2) 各主題領域における掲載先上位 10 タイトルと掲載数を確認した（第 7 表、第 8 表）。

第 7 表は、arXiv 全体で掲載数の多いタイトルを示したものである。journal-ref が付与された各論文・記事の掲載先を集計し、総掲載数が多いタイトルから順に、全体の 8 割までを占める上位 56 タイトルについて、総掲載数、全体におけるシェア、累積シェア、出版者を示した。journal-ref に記述されたタイトルは可能な限り名寄せを行っているが、journal-ref は基本的に著者によって記述され表記が統一されておらず、正式タイトルを識別できないものもあるため、arXiv 全体で掲載先の正確な異なりタイトル数は不明である（約 6,700 異なりタイトル数を確認している）。出版者の確認は ulrichweb を用い、確認時のものを記した。

総掲載数が多いのは、物理学分野と天文学における著名なタイトルで占められており、上位 10 タイトルで全体の 5 割を占めている。特に、素粒子、場の理論、重力、宇宙論を扱う *Physical Review D*、原子核、素粒子を扱う *Physics Letters B*、物理学全領域を扱う *Physical Review Letters*、

第 5 表 各主題領域における登録数と

	宇宙物理学			物性物理学			数学			高エネルギー物理学現象		
	登録数	J-ref	%	登録数	J-ref	%	登録数	J-ref	%	登録数	J-ref	%
1991	0	0	0.0	0	0	0.0	0	0	0.0	0	0	0.0
1992	59	42	71.2	219	81	37.0	277	114	41.2	755	574	76.0
1993	489	272	55.6	687	160	23.3	221	92	41.6	1,762	1,296	73.6
1994	1,029	436	42.4	1,323	284	21.5	319	94	29.5	2,502	1,826	73.0
1995	1,658	663	40.0	1,823	538	29.5	332	82	24.7	2,992	2,088	69.8
1996	2,418	660	27.3	2,545	902	35.4	325	38	11.7	3,314	2,301	69.4
1997	3,576	822	23.0	3,534	1,327	37.5	417	59	14.1	3,615	2,475	68.5
1998	4,746	1,165	24.5	4,636	1,967	42.4	2,077	511	24.6	3,806	2,646	69.5
1999	5,638	1,539	27.3	5,488	2,310	42.1	2,471	661	26.8	4,075	2,892	71.0
2000	6,341	2,168	34.2	6,023	2,517	41.8	3,016	761	25.2	4,124	2,889	70.1
2001	6,806	2,347	34.5	7,024	3,228	46.0	3,259	847	26.0	4,223	3,043	72.1
2002	7,027	3,797	54.0	7,762	3,684	47.5	4,289	1,127	26.3	4,243	2,990	70.5
2003	7,899	5,354	67.8	8,367	3,960	47.3	5,275	1,342	25.4	3,963	2,806	70.8
2004	8,242	4,974	60.3	8,953	4,382	48.9	6,693	1,616	24.1	4,136	2,792	67.5
2005	8,746	4,939	56.5	9,233	4,661	50.5	7,911	1,584	20.0	3,915	2,716	69.4
2006	9,268	4,460	48.1	9,243	4,488	48.6	9,724	1,531	15.7	3,948	2,112	53.5
2007	10,110	1,088	10.8	9,421	3,178	33.7	11,408	1,037	9.1	3,991	427	10.7
合計	84,052	34,726	41.3	86,281	37,667	43.7	58,014	11,496	19.8	55,364	35,873	64.8

	計算機科学			高エネルギー物理学格子論			高エネルギー物理学実験			数理論理学		
	登録数	J-ref	%	登録数	J-ref	%	登録数	J-ref	%	登録数	J-ref	%
1991	0	0	0.0	4	4	100.0	0	0	0.0	0	0	0.0
1992	0	0	0.0	322	276	85.7	0	0	0.0	0	0	0.0
1993	23	23	100.0	374	312	83.4	0	0	0.0	0	0	0.0
1994	238	102	42.9	446	378	84.8	67	36	53.7	0	0	0.0
1995	254	97	38.2	436	368	84.4	158	108	68.4	0	0	0.0
1996	232	100	43.1	558	470	84.2	166	93	56.0	0	0	0.0
1997	185	99	53.5	548	492	89.8	325	221	68.0	0	0	0.0
1998	304	137	45.1	623	542	87.0	407	268	65.8	232	107	46.1
1999	277	87	31.4	588	496	84.4	667	386	57.9	407	171	42.0
2000	445	146	32.8	508	412	81.1	736	481	65.4	521	250	48.0
2001	539	176	32.7	575	517	89.9	817	466	57.0	489	200	40.9
2002	578	191	33.0	583	270	46.3	798	449	56.3	702	348	49.6
2003	707	154	21.8	575	265	46.1	771	428	55.5	747	336	45.0
2004	856	283	33.1	586	310	52.9	885	464	52.4	874	416	47.6
2005	1,200	226	18.8	663	588	88.7	854	445	52.1	951	384	40.4
2006	1,676	353	21.1	504	279	55.4	848	352	41.5	937	309	33.0
2007	2,391	532	22.3	606	241	39.8	863	71	8.2	1,163	167	14.4
合計	9,906	2,363	23.9	8,499	6,220	73.2	8,362	4,268	51.0	7,023	2,688	38.3

注：高エネルギー物理学 4 領域という主題領域は，arXiv には設置されていないが，第 2 図の実数値を示すものとし

天文学分野のコアジャーナルである *Astrophysical Journal*，物性物理学や材料物理学を扱う *Physical Review B*，の 5 タイトルは総掲載数が 1 万件を超えている。これらのタイトルの総掲載数が多いのは，第 4 表に示した論文・記事の登録件数自体が多い主題領域を反映し，高エネルギー物

理学（素粒子），物性物理学，宇宙物理学関連の論文が相対的に多く登録されていると考えられる。出版者も物理学・天文学における代表的な学協会，研究所および著名な国際商業出版社が占めている。論文数のシェアでは，米国物理学会とエルゼビアが刊行するタイトルで全体の 5 割弱

journal-ref 付与数および付与率(1991～2007)

高エネルギー 物理学理論			量子物理学			物理学一般			一般相対性理論・ 宇宙論			原子核物理学理論		
登録数	J-ref	%	登録数	J-ref	%	登録数	J-ref	%	登録数	J-ref	%	登録数	J-ref	%
302	235	77.8	0	0	0.0	0	0	0.0	0	0	0.0	0	0	0.0
1,384	1,101	79.6	0	0	0.0	0	0	0.0	101	74	73.3	49	26	53.1
2,091	1,522	72.8	0	0	0.0	0	0	0.0	421	294	69.8	302	190	62.9
2,441	1,694	69.4	12	6	50.0	45	13	29.0	682	494	72.4	390	262	67.2
2,346	1,787	76.2	332	141	42.5	110	24	22.0	738	518	70.2	517	369	71.4
2,626	2,023	77.0	465	186	40.0	202	79	39.0	836	630	75.4	652	504	77.3
2,686	2,115	78.7	689	337	48.9	445	187	42.0	1,019	706	69.3	806	597	74.1
2,774	2,139	77.1	1,021	561	54.9	589	222	38.0	1,139	837	73.5	977	727	74.4
2,826	2,253	79.7	1,275	683	53.6	743	242	33.0	1,247	919	73.7	1,035	787	76.0
3,145	2,431	77.3	1,515	682	45.0	1,143	531	46.0	1,252	932	74.4	973	744	76.5
3,182	2,538	79.8	1,910	916	48.0	1,224	484	40.0	1,285	926	72.1	941	731	77.7
3,333	2,651	79.5	2,175	1,097	50.4	1,223	404	33.0	1,312	980	74.7	1,046	807	77.2
3,276	2,631	80.3	2,438	1,266	51.9	1,619	572	35.0	1,420	1,048	73.8	1,156	883	76.4
3,357	2,605	77.6	2,604	1,388	53.3	2,086	717	34.0	1,554	1,125	72.4	1,142	872	76.4
3,239	2,579	79.6	2,857	1,434	50.2	2,759	965	35.0	1,661	1,230	74.1	1,064	774	72.7
3,291	2,020	61.4	2,953	1,336	45.2	3,155	975	31.0	1,741	970	55.7	1,049	611	58.2
3,416	525	15.4	3,019	729	24.1	3,634	793	22.0	1,905	410	21.5	1,113	177	15.9
45,715	32,849	71.9	23,265	10,762	46.3	18,977	6,208	33.0	18,313	12,093	66.0	13,212	9,061	68.6
非線形科学			原子核物理学実験			数量的物理学			統計学			高エネルギー 物理学 4 領域 (参考)		
登録数	J-ref	%	登録数	J-ref	%	登録数	J-ref	%	登録数	J-ref	%	登録数	J-ref	%
0	0	0.0	0	0	0.0	0	0	0.0	0	0	0.0	306	239	78.1
0	0	0.0	0	0	0.0	0	0	0.0	0	0	0.0	2,461	1,951	79.3
0	0	0.0	0	0	0.0	0	0	0.0	0	0	0.0	4,227	3,130	74.0
0	0	0.0	4	3	75.0	0	0	0.0	0	0	0.0	5,456	3,934	72.1
0	0	0.0	24	17	70.8	0	0	0.0	0	0	0.0	5,932	4,351	73.3
0	0	0.0	57	35	61.4	0	0	0.0	0	0	0.0	6,664	4,887	73.3
0	0	0.0	69	50	72.5	0	0	0.0	0	0	0.0	7,174	5,303	73.9
2	2	100.0	110	88	80.0	0	0	0.0	0	0	0.0	7,610	5,595	73.5
4	4	100.0	174	131	75.3	0	0	0.0	0	0	0.0	8,156	6,027	73.9
670	270	40.3	174	141	81.0	0	0	0.0	0	0	0.0	8,513	6,213	73.0
676	254	37.6	245	187	76.3	0	0	0.0	0	0	0.0	8,797	6,564	74.6
735	272	37.0	295	231	78.3	0	0	0.0	0	0	0.0	8,957	6,360	71.0
715	259	36.2	323	230	71.2	163	50	30.7	0	0	0.0	8,585	6,130	71.4
760	286	37.6	474	357	75.3	506	178	35.2	0	0	0.0	8,964	6,171	68.8
766	292	38.1	460	340	73.9	546	177	32.4	0	0	0.0	8,671	6,328	73.0
835	271	32.5	441	239	54.2	591	185	31.3	0	0	0.0	8,591	4,763	55.4
734	113	15.4	555	77	13.9	715	184	25.7	265	109	41.1	8,876	1,264	14.2
5,897	2,023	34.3	3,405	2,126	62.4	2,521	774	30.7	265	109	41.1	117,940	79,210	67.2

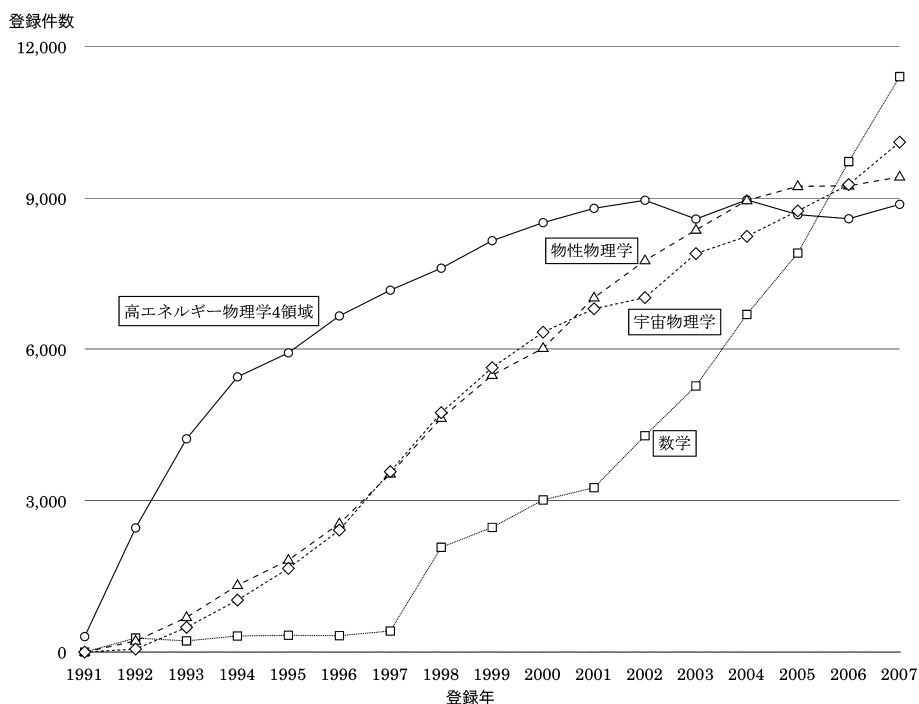
て便宜上加えている。

(46.5%) になる。

第 8 表は、主題領域ごとに集計した掲載先について、掲載数の多い順に上位 10 タイトルを示したものである。当然ながら、各主題領域に対応したタイトルがそれぞれ上位を占めている。宇宙物理学では天文学分野のコアジャーナルが、数学で

は統計学、確率論、幾何学関連のタイトルが、生物学では理論生物学、生物物理学のタイトルが含まれている。上位 10 タイトルのうち、アスタリスクが付与されたもの以外は全て学術雑誌であるが、計算機科学の場合、同分野の代表的成果公表メディアである会議録やシリーズものが他の主題

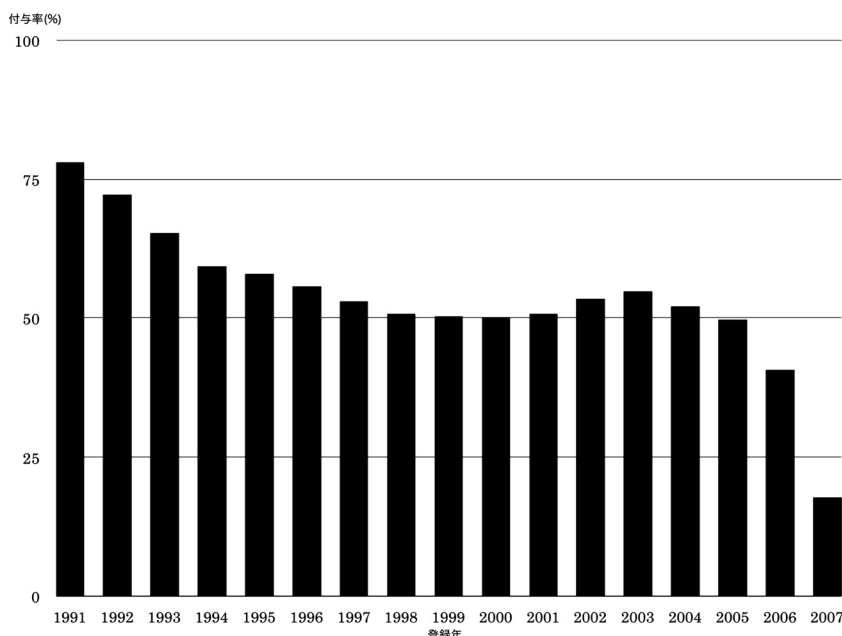
学術情報メディアとしての arXiv の位置づけ



第2図 登録件数上位4主題領域における登録件数の変遷

第6表 journal-ref の付与率

		1991～2005 (平均)	1991～2007 (平均)
1	高エネルギー物理学理論	77.7%	71.9%
2	高エネルギー物理学格子論	77.1%	73.2%
3	原子核物理学実験	75.1%	62.4%
4	原子核物理学理論	74.9%	68.6%
5	一般相対性理論・量子宇宙論	73.0%	66.0%
6	高エネルギー物理学現象	70.3%	64.8%
7	高エネルギー物理学実験	57.8%	51.0%
8	量子物理学	50.3%	46.3%
9	宇宙物理学	45.0%	41.2%
10	数理物理学	44.9%	38.3%
11	物性物理学	44.4%	43.7%
12	統計学	41.4%	41.1%
13	物理学一般	36.4%	32.7%
14	非線形科学	36.0%	34.3%
15	数量的生物学	33.3%	30.7%
16	計算機科学	31.2%	23.9%
17	数学	24.2%	19.8%
	全体	52.7%	47.1%



第3図 journal-ref の付与率

領域と比較して多く含まれている。掲載数上位 10 タイトルが各主題領域でどれだけの割合を占めるのかを示したのが累積シェアであり、高エネルギー物理学 4 領域、宇宙物理学、一般相対性理論・量子宇宙論、原子核物理学理論および実験においては上位 10 タイトルのみで 8 割前後を占めている。一方で、計算機科学、数学、数量的生物学、物理学一般は 2 割から 4 割弱を占めるにすぎない。上位 10 タイトルは、計算機科学以外はほぼすべてが学術雑誌であることから、この割合が高くかつ journal-ref 付与率が高い高エネルギー物理学 4 領域や原子核物理学といった主題領域においては、登録されている論文・記事は大部分が「学術雑誌」に掲載されたと考えられる。

3. 学術雑誌掲載論文網羅率

前節で示したように、journal-ref が付与された論文・記事の掲載先は学術雑誌が大半であったが、学術雑誌に掲載される論文を arXiv に登録することがどれだけ普及しているかを推定するために、arXiv に登録されかつ学術雑誌に掲載された論文が、どれだけあるタイトルの総掲載論文を網

羅しているかを調査した。タイトル X の A 年における arXiv 登録論文の総掲載論文網羅率は、以下の式によって算出した。

総掲載論文網羅率＝

$$\frac{\text{X 誌に A 年に掲載された論文のうち arXiv への総登録件数}}{\text{X 誌の A 年における総掲載論文}} \times 100$$

なお、対象期間は、1991 年から 2007 年であり、それ以外の期間に学術雑誌などに掲載された arXiv 登録論文は対象外とした。分子の数は、arXiv に登録された年ではなく、journal-ref に記載された掲載年をもとに X 誌の A 年における総登録数を集計している。次節で示すように、ある年に arXiv に登録された論文が同じ年に学術雑誌に掲載されるとは限らないためである。ただし、arXiv と出版元データベースとの間で論文のタイトル照合などの確認作業は行わず、あくまで arXiv における登録数と学術雑誌における掲載論文数とを比較しているため、網羅率を推測するにとどまる。

第 7 表に示したタイトルについて、1991 年から 2005 年と 1991 年から 2007 年における平均

学術情報メディアとしての arXiv の位置づけ

第 7 表 arXiv 全体における掲載数上位タイトル (全体の 80%まで)

順位	タイトル名	総掲載数	シェア	累積シェア	出版者
1	Phys. Rev. D.	24,379	11.5%	11.5%	米国物理学会
2	Phys. Lett. B.	15,666	7.4%	19.0%	エルゼビア
3	Phys. Rev. Lett.	12,484	5.9%	24.9%	米国物理学会
4	Astrophys. J.	11,810	5.6%	30.5%	米国天文学会
5	Phys. Rev. B.	10,183	4.8%	35.3%	米国物理学会
6	Nucl. Phys. B.	8,599	4.1%	39.4%	エルゼビア
7	MNRAS	6,748	3.2%	42.6%	王立天文協会
8	JHEP	6,533	3.1%	45.7%	イタリア国際先端研究所
9	Phys. Rev. A.	4,895	2.3%	48.0%	米国物理学会
10	Phys. Rev. C.	4,729	2.2%	50.2%	米国物理学会
11	Phys. Rev. E.	4,631	2.2%	52.4%	米国物理学会
12	Astron. Astrophys.	4,203	2.0%	54.4%	EDP Sciences
13	J. Phys. A.	3,929	1.9%	56.3%	英国物理学会
14	Class. Quant. Grav.	3,878	1.8%	58.1%	英国物理学会
15	*Nucl. Phys. B. Proc. Suppl.	3,719	1.8%	59.9%	エルゼビア
16	Nucl. Phys. A.	3,053	1.4%	61.3%	エルゼビア
17	Int. J. Mod. Phys. A.	2,997	1.4%	62.7%	World Scientific Publishing
18	Mod. Phys. Lett. A.	2,895	1.4%	64.1%	World Scientific Publishing
19	Eur. Phys. J. C.	2,591	1.2%	65.3%	EDP Sciences
20	J. Math. Phys.	2,020	1.0%	66.3%	米国物理学協会
21	Phys. Lett. A.	1,923	0.9%	67.2%	エルゼビア
22	Eur. Phys. Lett.	1,569	0.7%	68.0%	EDP Sciences
23	J. Phys. G.	1,420	0.7%	68.6%	英国物理学会
24	Prog. Theor. Phys.	1,249	0.6%	69.2%	京都大学基礎物理学研究所, 日本物理学会
25	Physica A.	1,154	0.5%	69.8%	エルゼビア
26	Commun. Math. Phys.	1,128	0.5%	70.3%	シュプリンガー
27	Acta Phys. Pol. B	1,124	0.5%	70.8%	ヤギェウォ大学
28	Nucl. Instrum. Meth. A.	1,013	0.5%	71.3%	エルゼビア
29	J. Phys. CM	992	0.5%	71.8%	英国物理学会
30	*AIP Conf. Proc.	979	0.5%	72.2%	米国物理学協会
31	Eur. Phys. J. B.	946	0.4%	72.7%	EDP Sciences
32	Gen. Rel. Grav.	912	0.4%	73.1%	シュプリンガー
33	J. Phys. Soc. Jpn.	840	0.4%	73.5%	日本物理学会
34	Int. J. Mod. Phys. D.	806	0.4%	73.9%	World Scientific Publishing
35	Ann. Phys. (N.Y.)	802	0.4%	74.3%	エルゼビア
36	JCAP	754	0.4%	74.6%	イタリア国際先端研究所, 英国物理学会
37	Eur. Phys. J. A.	746	0.4%	75.0%	EDP Sciences
38	*eConf	707	0.3%	75.3%	スタンフォード線形加速器センター
39	Astron. J.	702	0.3%	75.7%	米国天文学会
40	J. Chem. Phys.	699	0.3%	76.0%	米国物理学協会
41	Appl. Phys. Lett.	685	0.3%	76.3%	米国物理学協会
42	Astropart. Phys.	637	0.3%	76.6%	エルゼビア
43	Phys. Atom. Nucl.	624	0.3%	76.9%	シュプリンガー
44	Z. Phys. C	622	0.3%	77.2%	シュプリンガー
45	Astrophys. Space Sci	597	0.3%	77.5%	シュプリンガー
46	J. Stat. Phys.	568	0.3%	77.8%	シュプリンガー
47	Int. J. Theor. Phys.	561	0.3%	78.0%	シュプリンガー
48	*PoS LAT	544	0.3%	78.3%	イタリア国際先端研究所
49	JETP Lett.	541	0.3%	78.5%	シュプリンガー
50	New J. Phys.	502	0.2%	78.8%	英国物理学会
51	J. Stat. Mech.	463	0.2%	79.0%	英国物理学会
52	Lett. Math. Phys.	453	0.2%	79.2%	シュプリンガー
53	nature	440	0.2%	79.4%	Nature Publishing Group
54	Ann. Statist.	405	0.2%	79.6%	数理統計学会
55	Int. J. Mod. Phys. B.	396	0.2%	79.8%	World Scientific Publishing
56	New Astron. Rev.	381	0.2%	80.0%	エルゼビア
合計		168,826			

注: journal-ref が付与された論文・記事は合計 211,085 件。アスタリスクがついているタイトルは、会議録。

第 8 表 各主題領域における掲載数上位 10 タイトル

物性物理学					宇宙物理学				
順位	タイトル名	掲載数	シェア	累積シェア	順位	タイトル名	掲載数	シェア	累積シェア
1	Phys. Rev. B.	9,957	26.5%	26.5%	1	Astrophys. J.	11,748	33.8%	33.8%
2	Phys. Rev. Lett.	6,007	16.0%	42.5%	2	MNRAS	6,690	19.3%	53.1%
3	Phys. Rev. E.	3,380	9.0%	51.4%	3	Astron. Astrophys.	4,171	12.0%	65.1%
4	J. Phys. A.	1,141	3.0%	54.5%	4	Phys. Rev. D.	2,271	6.5%	71.6%
5	Phys. Rev. A.	1,121	3.0%	57.5%	5	Astron. J.	702	2.0%	73.7%
6	Eur. Phys. Lett.	1,112	3.0%	60.4%	6	Phys. Rev. Lett.	581	1.7%	75.3%
7	J. Phys. CM	956	2.5%	63.0%	7	Astrophys. Space Sci.	526	1.5%	76.9%
8	Eur. Phys. J. B.	835	2.2%	65.2%	8	Astropart. Phys.	458	1.3%	78.2%
9	Physica A.	807	2.1%	67.3%	9	Phys. Lett. B.	378	1.1%	79.3%
10	J. Phys. Soc. Jpn.	777	2.1%	69.4%	10	New Astron. Rev.	377	1.1%	80.4%
合計		26,093			合計		27,902		
数学					高エネルギー物理学現象				
順位	タイトル名	掲載数	シェア	累積シェア	順位	タイトル名	掲載数	シェア	累積シェア
1	Ann. Statist.	401	3.6%	3.6%	1	Phys. Rev. D.	10,946	30.5%	30.5%
2	Ann. Probab.	333	3.0%	6.5%	2	Phys. Lett. B.	7,165	20.0%	50.5%
3	Algebr. Geom. Topol.	277	2.5%	9.0%	3	Nucl. Phys. B.	2,312	6.4%	56.9%
4	Ann. Appl. Prob.	267	2.4%	11.4%	4	Eur. Phys. J. C.	1,661	4.6%	61.6%
5	Ann. Math.	248	2.2%	13.6%	5	JHEP	1,466	4.1%	65.6%
6	Geom. Topol.	244	2.2%	15.8%	6	Phys. Rev. Lett.	1,345	3.7%	69.4%
7	J. Algebra	223	2.0%	17.7%	7	*Nucl. Phys. B. Proc. Suppl.	1,042	2.9%	72.3%
8	Commun. Math. Phys.	218	1.9%	19.7%	8	Nucl. Phys. A.	934	2.6%	74.9%
9	*Proceedings of the ICM	164	1.5%	21.1%	9	Int. J. Mod. Phys. A.	826	2.3%	77.2%
10	Int. Math. Res. Not.	159	1.4%	22.6%	10	Mod. Phys. Lett. A.	756	2.1%	79.3%
合計		2,534			合計		28,453		
高エネルギー物理学理論					量子物理学				
順位	タイトル名	掲載数	シェア	累積シェア	順位	タイトル名	掲載数	シェア	累積シェア
1	Phys. Rev. D.	5,278	16.1%	16.1%	1	Phys. Rev. A.	3,164	29.4%	29.4%
2	Phys. Lett. B.	5,071	15.4%	31.5%	2	Phys. Rev. Lett.	1,286	11.9%	41.3%
3	Nucl. Phys. B.	5,038	15.3%	46.8%	3	J. Phys. A.	819	7.6%	48.9%
4	JHEP	4,749	14.5%	61.3%	4	Phys. Lett. A.	664	6.2%	55.1%
5	Int. J. Mod. Phys. A.	1,477	4.5%	65.8%	5	J. Math. Phys.	228	2.1%	57.2%
6	Mod. Phys. Lett. A.	1,454	4.4%	70.2%	6	J. Opt. B.	194	1.8%	59.0%
7	Class. Quant. Grav.	1,077	3.3%	73.5%	7	New J. Phys.	157	1.5%	60.5%
8	J. Phys. A.	968	2.9%	76.4%	8	Phys. Rev. E.	149	1.4%	61.9%
9	J. Math. Phys.	799	2.4%	78.9%	9	Found. Phys.	149	1.4%	63.3%
10	Phys. Rev. Lett.	586	1.8%	80.7%	10	Quant. Inf. & Comp.	147	1.4%	64.6%
合計		26,497			合計		6,957		

網羅率が高い順に示したのが、第 9 表である。先に示したように、journal-ref の付与率は 2006 および 2007 年において著しく減少しているため、ここではより実態を正確に表していると考えられる 2005 年までの期間を優先して網羅率を並べている。1991 年から 2005 年の場合、*Journal of High Energy Physics* (JHEP) の 99.6% から、*nature* の 0.7% まで、タイトルによって網羅率に

は大きな開きが見られる。総登録件数と網羅率との間の相関係数は 0.34 であることから、両者の間には強い相関はない。*JHEP* と同様、*Journal of Cosmology and Astroparticle Physics* (JCAP) も、イタリア国際先端研究所と英国物理学会が刊行し電子ジャーナルでも提供されているが、この 2 誌は、journal-ref の未付与等の理由から期間全体では網羅率は 100% ではないが、網羅率が 100%

学術情報メディアとしての arXiv の位置づけ

第 8 表 つづき

物理学一般

順位	タイトル名	掲載数	シェア	累積シェア
1	Phys. Rev. A.	426	6.9%	6.9%
2	Phys. Rev. E.	412	6.6%	13.5%
3	Phys. Rev. Lett.	371	6.0%	19.5%
4	Nucl. Instrum. Meth. A.	350	5.6%	25.1%
5	*eConf	333	5.4%	30.5%
6	Physica A.	145	2.3%	32.8%
7	J. Chem. Phys.	122	2.0%	34.8%
8	Phys. Lett. A.	114	1.8%	36.6%
9	J. Phys. B.	106	1.7%	38.3%
10	Opt. Lett.	98	1.6%	39.9%
合計		2,477		

原子核物理学理論

順位	タイトル名	掲載数	シェア	累積シェア
1	Phys. Rev. C.	3,453	38.1%	38.1%
2	Nucl. Phys. A.	1,561	17.2%	55.3%
3	Phys. Lett. B.	900	9.9%	65.3%
4	Phys. Rev. Lett.	427	4.7%	70.0%
5	Eur. Phys. J. A.	369	4.1%	74.1%
6	J. Phys. G.	355	3.9%	78.0%
7	Acta Phys. Pol. B	129	1.4%	79.4%
8	Phys. Rev. D.	122	1.3%	80.7%
9	Prog. Theor. Phys.	117	1.3%	82.0%
10	Int. J. Mod. Phys. E.	107	1.2%	83.2%
合計		7,540		

高エネルギー物理学格子論

順位	タイトル名	掲載数	シェア	累積シェア
1	*Nucl. Phys. B. Proc. Suppl.	1,997	32.1%	32.1%
2	Phys. Rev. D.	1,297	20.9%	53.0%
3	Phys. Lett. B.	664	10.7%	63.6%
4	Nucl. Phys. B.	610	9.8%	73.4%
5	*PoS LAT	544	8.7%	82.2%
6	JHEP	217	3.5%	85.7%
7	Phys. Rev. Lett.	140	2.3%	87.9%
8	Prog. Theor. Phys.	75	1.2%	89.1%
9	Nucl. Phys. A.	68	1.1%	90.2%
10	Int. J. Mod. Phys. A.	60	1.0%	91.2%
合計		5,672		

一般相対性理論・量子宇宙論

順位	タイトル名	掲載数	シェア	累積シェア
1	Phys. Rev. D.	3,676	30.4%	30.4%
2	Class. Quant. Grav.	2,584	21.4%	51.8%
3	Gen. Rel. Grav.	732	6.1%	57.8%
4	Int. J. Mod. Phys. D.	467	3.9%	61.7%
5	Phys. Lett. B.	418	3.5%	65.1%
6	J. Math. Phys.	381	3.2%	68.3%
7	Mod. Phys. Lett. A.	323	2.7%	71.0%
8	Phys. Rev. Lett.	321	2.7%	73.6%
9	Phys. Lett. A.	207	1.7%	75.3%
10	Grav. Cosmol.	187	1.5%	76.9%
合計		9,296		

計算機科学

順位	タイトル名	掲載数	シェア	累積シェア
1	J. Artificial Intelligence Research	110	4.5%	4.5%
2	LMCS	76	3.1%	7.5%
3	*Lect. Notes. Comput. Sci.	62	2.5%	10.0%
4	*Symposium on Design, Test, Integration and Packaging of MEMS/ MOEMS	49	2.0%	12.0%
5	*Proc. Annual Meeting of the Association for Computational Linguistics	43	1.7%	13.8%
6	*Proc. COLING	42	1.7%	15.5%
7	IEEE Trans. Inform. Theory	40	1.6%	17.1%
8	Theory and Practice of Logic Programming	34	1.4%	18.5%
9	*eConf	31	1.3%	19.7%
10	Int. J. Advanced Robotic Systems	29	1.2%	20.9%
合計		516		

高エネルギー物理学実験

順位	タイトル名	掲載数	シェア	累積シェア
1	Phys. Rev. Lett.	857	20.1%	20.1%
2	Phys. Lett. B.	823	19.3%	39.4%
3	Phys. Rev. D.	686	16.1%	55.4%
4	Eur. Phys. J. C.	462	10.8%	66.3%
5	Nucl. Instrum. Meth. A.	277	6.5%	72.8%
6	*Nucl. Phys. B. Proc. Suppl.	179	4.2%	76.9%
7	Int. J. Mod. Phys. A.	167	3.9%	80.9%
8	*eConf	125	2.9%	83.8%
9	*AIP Conf. Proc.	53	1.2%	85.0%
10	Nucl. Phys. A.	52	1.2%	86.2%
合計		3,681		

第 8 表 つづき

数理論理学

順位	タイトル名	掲載数	シェア	累積シェア
1	J. Phys. A.	461	17.2%	17.2%
2	J. Math. Phys.	313	11.6%	28.8%
3	Commun. Math. Phys.	175	6.5%	35.3%
4	Lett. Math. Phys.	72	2.7%	38.0%
5	J. Stat. Phys.	69	2.6%	40.6%
6	SIGMA	65	2.4%	43.0%
7	J. Nonlin. Math. Phys.	57	2.1%	45.1%
8	Phys. Lett. A.	56	2.1%	47.2%
9	Rev. Math. Phys.	53	2.0%	49.1%
10	J. Geom. Phys.	43	1.6%	50.7%
合計		1,364		

原子核物理学実験

順位	タイトル名	掲載数	シェア	累積シェア
1	Phys. Rev. C.	489	23.0%	23.0%
2	Phys. Rev. Lett.	340	16.0%	39.0%
3	Nucl. Phys. A.	326	15.3%	54.4%
4	J. Phys. G.	178	8.4%	62.8%
5	Phys. Lett. B.	172	8.1%	70.9%
6	Nucl. Instrum. Meth. A.	114	5.4%	76.2%
7	Eur. Phys. J. A.	92	4.3%	80.6%
8	Acta Phys. Hung. A	42	2.0%	82.5%
9	Eur. Phys. J. C.	37	1.7%	84.3%
10	Acta Phys. Pol. B	30	1.4%	85.7%
合計		1,820		

統計学

順位	タイトル名	掲載数	シェア	累積シェア
1	Statistical Science	49	45.0%	45.0%
2	Ann. Appl. Stat.	28	25.7%	70.6%
3	*IMS Lecture Notes Monograph Series	13	11.9%	82.6%
4	Elec. J. Stat.	6	5.5%	88.1%
5	Ann. Statist.	3	2.8%	90.8%
6	Comput. Methods Programs Biomed.	2	1.8%	92.7%
7	その他 (掲載数が 1 件の みのタイトル)	8	7.3%	100.0%
合計		109		

注: アスタリスクがついたタイトルは、会議録あるいはモノグラフシリーズ。

非線形科学

順位	タイトル名	掲載数	シェア	累積シェア
1	Phys. Rev. E.	412	20.4%	20.4%
2	J. Phys. A.	206	10.2%	30.5%
3	Phys. Rev. Lett.	162	8.0%	38.6%
4	Phys. Lett. A.	86	4.3%	42.8%
5	Physica D.	77	3.8%	46.6%
6	J. Nonlin. Math. Phys.	68	3.4%	50.0%
7	SIGMA	58	2.9%	52.8%
8	J. Math. Phys.	43	2.1%	55.0%
9	Eur. Phys. Lett.	40	2.0%	56.9%
10	Physica A.	39	1.9%	58.9%
合計		1,191		

数量的生物学

順位	タイトル名	掲載数	シェア	累積シェア
1	Phys. Rev. E.	82	10.6%	10.6%
2	Phys. Rev. Lett.	46	5.9%	16.5%
3	PNAS	39	5.0%	21.6%
4	J. Theor. Biol.	36	4.7%	26.2%
5	Biophys. J.	24	3.1%	29.3%
6	Physica A.	21	2.7%	32.0%
7	BMC Bioinformatics	18	2.3%	34.4%
8	Phys. Biol.	15	1.9%	36.3%
9	J. Chem. Phys.	14	1.8%	38.1%
10	Bioinformatics	14	1.8%	39.9%
合計		309		

の年があること、投稿ファイルの登録先に arXiv を推奨していることから、arXiv に登録された論文でほぼすべての掲載論文を網羅していると考えられる。上位のタイトルは、素粒子、場の理論、重力、宇宙論を主に扱う学術雑誌が多いことから、arXiv はこれらの専門領域の研究者によって活発に利用されていると考えられる。

次に、arXiv 全体での登録件数上位 10 タイト

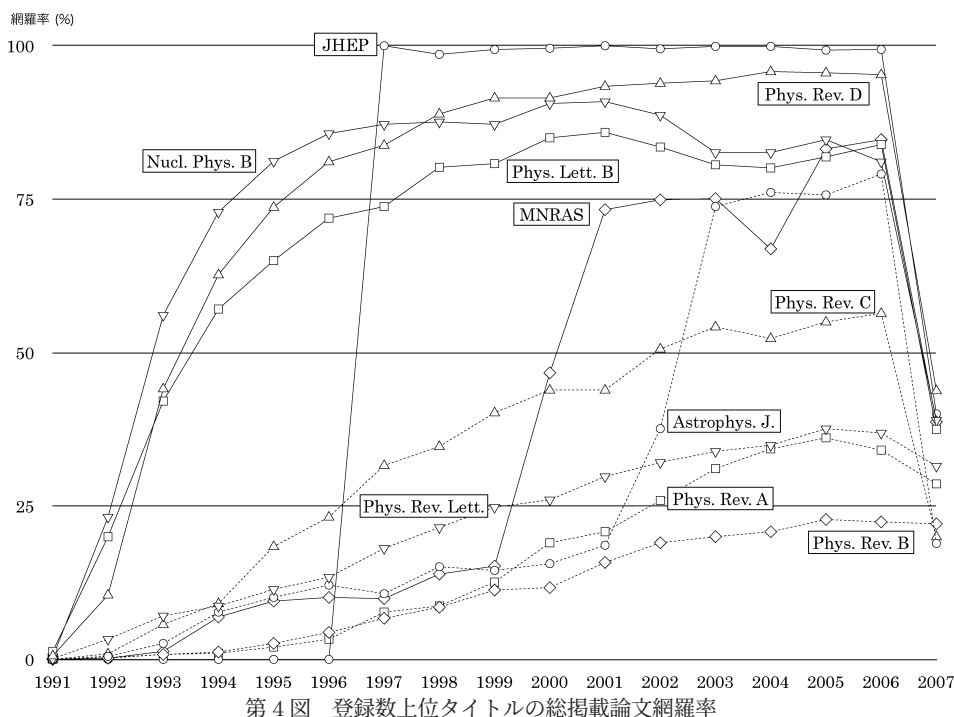
ルを対象に、年ごとにそれぞれのタイトルにおける論文の総掲載論文網羅率を示したのが、第 4 図である。2006 年時点において、*Journal of High Energy Physics* (JHEP) (99.4%) を筆頭に、*Physical Review D* (95.3%)、*Monthly Notices of the Royal Astronomical Society* (MNRAS) (84.7%)、*Physics Letters B* (83.9%)、*Nuclear Physics B* (81.1%)、*Astrophysical Journal* (79.1%) は非

学術情報メディアとしての arXiv の位置づけ

第 9 表 掲載数上位タイトルの学術雑誌掲載論文網羅率

タイトル名	1991～2005			1991～2007		
	網羅率	総登録件数	総掲載論文数	網羅率	総登録件数	総掲載論文数
JHEP	99.6%	5,012	5,031	89.4%	6,533	7,306
JCAP	99.0%	389	393	84.8%	754	889
Eur. Phys. J. C.	85.1%	2,200	2,586	79.9%	2,588	3,238
Phys. Rev. D.	79.4%	21,015	26,472	78.0%	24,374	31,266
*PoS LAT	78.1%	285	365	56.7%	549	969
Nucl. Phys. B.	74.5%	8,038	10,787	73.7%	8,585	11,653
J. Stat. Mech.	67.1%	171	255	64.1%	456	711
Phys. Lett. B.	63.4%	14,390	22,684	63.4%	15,653	24,694
Class. Quant. Grav.	60.5%	3,190	5,272	61.7%	3,872	6,271
Astropart. Phys.	58.5%	527	901	59.4%	636	1,071
Mod. Phys. Lett. A.	54.2%	2,620	4,837	53.6%	2,894	5,399
Int. J. Mod. Phys. D.	53.1%	665	1,253	49.3%	806	1,634
Int. J. Mod. Phys. A.	47.7%	2,594	5,439	47.4%	2,994	6,310
Ann. Phys. (N.Y.)	44.2%	676	1,531	45.8%	802	1,751
Prog. Theor. Phys.	41.7%	1,087	2,609	43.5%	1,249	2,874
Gen. Rel. Grav.	41.4%	782	1,888	42.1%	905	2,149
MNRAS	41.2%	5,174	12,570	45.1%	6,745	14,972
Eur. Phys. J. A.	35.5%	589	1,661	34.7%	746	2,151
Lett. Math. Phys.	35.3%	405	1,146	35.0%	447	1,278
*Nucl. Phys. B. Proc. Suppl.	34.9%	3,654	10,470	31.8%	3,719	11,696
J. Phys. G.	32.9%	1,203	3,652	33.5%	1,415	4,229
Phys. Rev. C.	32.0%	4,029	12,592	32.7%	4,729	14,464
New Astron. Rev.	31.1%	291	937	29.9%	381	1,275
Commun. Math. Phys.	30.5%	972	3,191	30.6%	1,124	3,674
Z. Phys. C.	29.4%	619	2,102	29.4%	619	2,102
New J. Phys.	29.1%	241	827	31.7%	497	1,567
Astrophys. J.	28.6%	9,066	31,645	31.8%	11,804	37,128
J. Math. Phys.	27.8%	1,734	6,231	28.5%	2,015	7,061
Acta Phys. Pol. B.	27.3%	970	3,553	27.7%	1,123	4,049
J. Phys. A.	25.8%	3,072	11,920	27.7%	3,916	14,124
Nucl. Phys. A.	24.3%	2,628	10,820	26.4%	3,053	11,564
Phys. Rev. Lett.	22.3%	9,720	43,574	24.1%	12,410	51,422
Eur. Phys. J. B.	21.6%	747	3,452	22.2%	943	4,241
*eConf	21.3%	703	3,298	21.4%	707	3,298
Astron. Astrophys.	18.0%	3,799	21,161	16.8%	4,200	25,047
Int. J. Theor. Phys.	16.7%	454	2,714	17.7%	558	3,153
Eur. Phys. Lett.	16.3%	1,182	7,235	19.9%	1,548	7,790
Phys. Atom. Nucl.	15.2%	556	3,659	15.0%	623	4,146
Phys. Rev. A.	14.5%	3,441	23,777	17.2%	4,867	28,340
J. Stat. Phys.	14.3%	455	3,175	16.2%	562	3,464
Phys. Rev. E.	14.3%	3,621	25,305	15.9%	4,613	28,975
Ann. Statist.	12.5%	205	1,645	21.6%	405	1,877
Phys. Lett. A.	11.9%	1,644	13,779	12.2%	1,911	15,724
Int. J. Mod. Phys. B.	11.0%	330	2,992	12.0%	395	3,286
Astrophys. Space Sci.	10.7%	550	5,123	11.0%	596	5,435
Phys. Rev. B.	10.3%	7,487	72,845	11.9%	10,087	84,553
Physica A.	10.3%	890	8,670	11.2%	1,139	10,127
JETP Lett.	9.6%	459	4,757	10.1%	536	5,320
Astron. J.	9.4%	610	6,458	9.6%	700	7,269
J. Phys. Soc. Jpn.	7.1%	652	9,191	8.1%	837	10,277
J. Phys. CM.	4.3%	697	16,288	5.1%	975	19,142
Nucl. Instrum. Meth. A.	4.3%	831	19,466	4.6%	1,013	21,831
*AIP Conf. Proc.	2.0%	784	38,679	2.3%	979	43,164
J. Chem. Phys.	1.4%	491	35,359	1.7%	692	40,699
Appl. Phys. Lett.	1.0%	397	39,442	1.3%	676	51,076
Nature	0.7%	331	45,645	0.9%	438	51,057

*が付与されたタイトルは、会議録。



常に高い網羅率になっている。特に *JHEP* は1997年の創刊からほぼ100%の網羅率を維持しており、先に述べた理由から、同誌に掲載された論文はほぼすべてarXivに登録されていると考えられる。天文学のタイトルである *MNRAS* と *Astrophysical Journal* は2000年前後になって急激に網羅率が上がっている。

登録件数が多くかつ網羅率が高いタイトルは天文学の2タイトルを除き、皆、素粒子を扱っており、その割合の上昇率から判断すると、高エネルギー物理学、特に掲載先上位タイトルを考慮すれば高エネルギー物理学理論や高エネルギー物理学現象といった領域では、arXivに論文を登録することがarXiv創設から4,5年で現在の程度まで普及したと考えられる。*Physical Review D* 以外の *Physical Review* 各誌についても、1990年代前半では網羅率はごくわずかであったが、その割合は年々上昇していることが読み取れる。依然として、物理学分野のコアジャーナルである *Physical Review* 各誌内においても、素粒子などを扱う *Physical Review D* とそれ以外では網羅率に大き

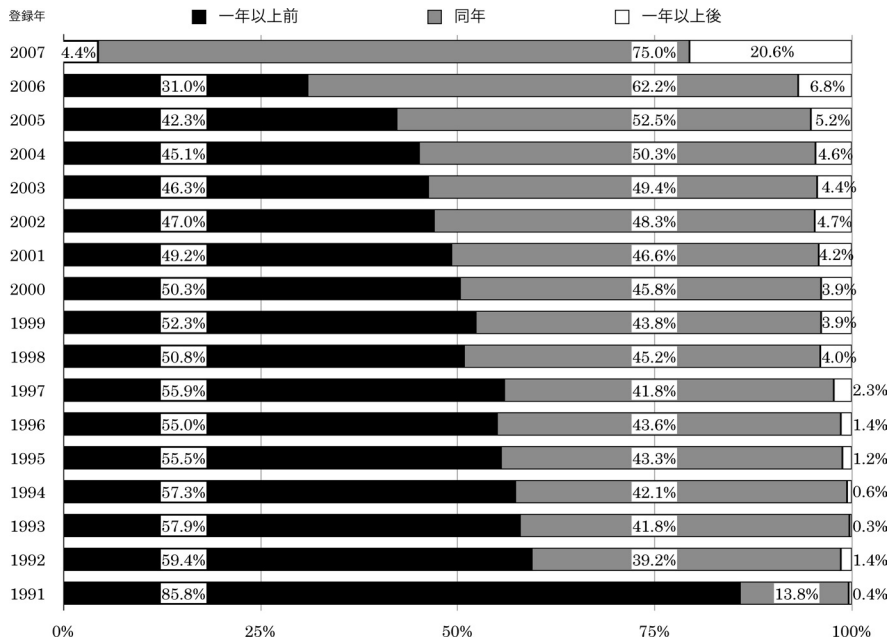
な開きがあるが、その差は縮まっている。

C. 登録年と掲載年の差（即時性）

前述したように、arXivの学術情報メディアの属性の一つとして即時性が指摘され、論文・記事が掲載先で公刊される以前にarXivに登録されている割合が高ければそれだけ、即時性という性質を満たしていると解釈することができる。実際に、journal-refの分析からも、掲載前の1から2年前にarXivに登録されていると推察できていることがわかっている。

ここでは即時性を、「arXivに登録された年と実際に学術雑誌等に掲載された年との差」とし、1) 登録年と掲載年に差がないもの、2) 登録年が掲載年よりも前のもの（早期アクセス）、3) 登録年が掲載年以降のもの（バックアクセス）、の3種類に分けた。その結果、arXiv全体では、掲載年と登録年の差がないものが46.6%、登録年が掲載年よりも前のものが42.3%を占めており、両者を合わせると9割近くになる。掲載年と登録年の差がないものは、可能性として学術雑誌等に掲載

学術情報メディアとしての arXiv の位置づけ



第 5 図 arXiv 全体における即時性の分布

前と掲載後に arXiv へ登録されたものの双方を含んでおり、より厳密には掲載前に arXiv に登録されたものだけが即時性の機能を果たしていることになるが、journal-ref が付与された論文・記事すべての正確な掲載年月日を知ることは不可能であるため、ここでは区別していない。それを考慮しても、arXiv に登録された論文・記事で学術雑誌等に掲載された大部分のものは、刊行前にアクセス可能な状態になっていると考えられる。

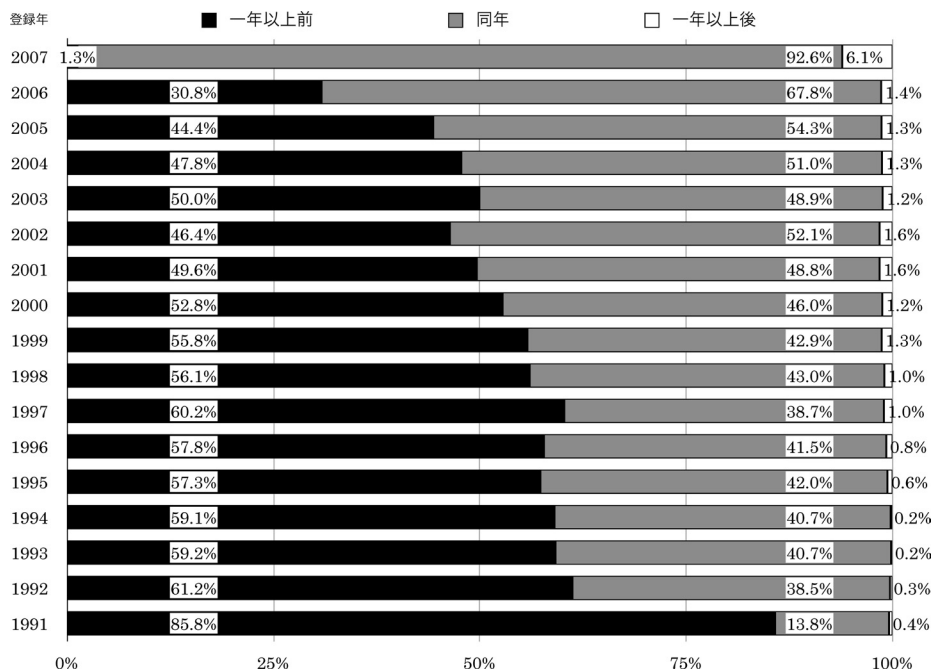
即時性の分布を論文・記事の登録年ごとに示したのが、第 5 図である。2007 年を除くすべての年で、同年と掲載年前年が 9 割以上を占めていることがわかる。2007 年は journal-ref 付与率が 17.8%であり、それが 2006 年以前とは異なる傾向をもたらしていると考えられる。経年変化に注目すると、1) 掲載年前年に登録される割合（早期アクセス）が減少している傾向と 2) 登録年と掲載年とに差がない割合と掲載年以降に登録される割合（バックアクセス）が上昇する傾向が見られることがわかる。1992 年当時は、掲載前年に arXiv に登録されるものが 59.4%であったが、2002 年には 47.0%にまで低下し同年（48.3%）

が占める割合のほうが高くなり、2005 年になってもその傾向は続いている。一方で、同年の割合は 1992 年の 39.2%から 2005 年の 52.5%に上昇し、掲載年以降に arXiv に登録される割合も依然として一桁台ではあるが、年々わずかながら上昇していることがわかる。過去に掲載された論文は、最大で 117 年前の論文（physics/9908003: *Comptes rendus de l'Académie des sciences* に 1882 年に掲載された論文を翻訳したもの）が登録されており、論文への早期アクセスというよりはむしろバックアクセスを提供することを目的として登録したものと考えられる。arXiv は本来、掲載前の学術雑誌論文への早期アクセスを提供することが目的であり、現在でもその機能を十分果たしていると考えられるが、それだけではなく過去に発表された論文を登録するアーカイブとしての機能をも一部ではあるが包含し始めていることを第 5 図から読み取ることができる。

次に、主題領域ごとに即時性を集計した結果が第 10 表であり、登録数および journal-ref 付与率と同様に、主題領域によって差が大きく、大きく以下の 3 パターンに分けることができる。

第10表 各主題領域における e-print の登録年と掲載年の差の割合（統計学を除く）

主題領域	掲載年前年	同年	掲載年以降	パターン
高エネルギー物理学格子論	60.3%	39.0%	0.7%	1
高エネルギー物理学現象	50.7%	48.2%	1.1%	1
原子核物理学実験	50.6%	46.8%	2.6%	1
高エネルギー物理学理論	50.6%	48.2%	1.2%	1
原子核物理学理論	49.9%	47.6%	2.5%	1
数学	47.8%	31.6%	20.6%	3
数理物理学	47.3%	46.2%	6.5%	1
一般相対性理論・量子宇宙論	46.7%	48.4%	4.9%	1
数理物理学	46.1%	39.4%	14.5%	3
高エネルギー物理学実験	42.5%	55.0%	2.5%	1
物性物理学	41.8%	52.9%	5.3%	1
非線形科学	39.9%	48.6%	11.4%	3
宇宙物理学	37.9%	59.8%	2.4%	1
物理学一般	32.8%	51.4%	15.8%	3
数論的生物学	29.5%	54.1%	16.4%	3
計算機科学	11.9%	50.3%	37.8%	2



第6図 高エネルギー物理学4領域全体における即時性の分布

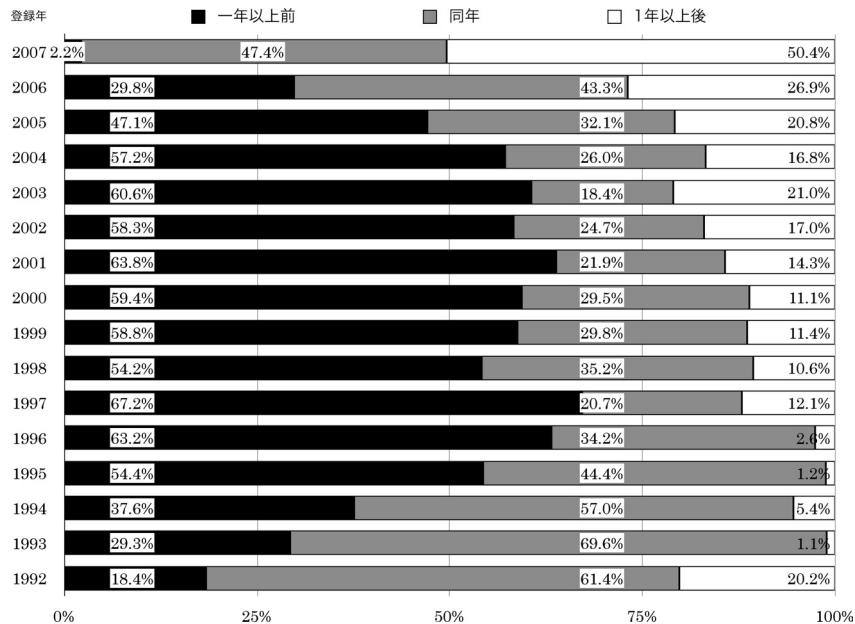
一つ目は、掲載年と登録年が同年のものと登録年が掲載年よりも前のものが9割以上のパターンであり、典型例である高エネルギー物理学の4領域は、その割合が100%に近い。

二つ目は、登録年が掲載年以降のものが占める割合が比較的高いパターンで、計算機科学では掲

載年以降に登録されたものが4割弱になっており、高エネルギー物理学などの第一のパターンに該当する主題領域とは異なる傾向を示している。

三つ目は、全体としては早期アクセスを提供しているが同時にバックアクセスをも提供しているパターンである。たとえば、数学では、掲載年以

学術情報メディアとしての arXiv の位置づけ



第 7 図 数学における即時性の分布

前が 47.8% であり全体で最も高い割合を占めるが、同時に掲載年以降に登録されたものも 20.6% を占めている。同様の傾向は、数量的生物学、物理学一般、数理論理学、非線形科学といった主題領域においても見られる。

主題領域ごとに経年変化を見ると、たとえば、高エネルギー物理学と数学ではかなり異なる分布を示していることがわかる。高エネルギー物理学は arXiv 全体と似た傾向、すなわち早期アクセスが大半を占めているが徐々に減少していく傾向を示している（第 6 図）。数学においては、早期アクセスの割合が増加から減少に転じ、同時にバックアクセスの割合が増加する傾向にあることがわかり、過去に発表された論文を登録するアーカイブとしても機能している（第 7 図）。

物理学系の主題領域では、現在でも arXiv は学術雑誌掲載前論文の早期アクセスを提供していると見なすことができる。より厳密に、「登録年が掲載年よりも早いもの」を早期アクセスを提供しているとしても、やはり高エネルギー物理学の 4 領域が上位を占めることがわかる。特に、高エネルギー物理学格子論は、60.4% が掲載年前年に登録

されている。しかし、計算機科学においては、早期アクセスの割合は約 12% にすぎず、arXiv が早期アクセスを提供しているとは言い難い。数学においては、掲載年前年の論文の割合は、他の主題領域と比較して大幅に低いわけではないが、早期アクセスと同時にバックアクセスも提供していると言え、異なるパターンを示している。

IV. 学術情報メディアとしての arXiv の位置づけ

本稿では、1991 年から 2007 年までに arXiv に登録されたすべての論文・記事の分析を通して、その学術情報メディアとしての位置づけを明らかにすることを目的とした。以下では、調査結果に基づいて、先に設定した三つの仮説を検討する。

本研究の結果をまとめると、以下の 4 点になる。

- 1) 登録件数は毎年増加傾向にあり、創設から発展し続けている。主題領域では依然として高エネルギー物理学 4 領域が総登録数では最多であるが、今後は宇宙物理学、物性物理学、

数学が年間登録数および総登録数でも中心となる可能性が高い。

- 2) journal-ref の付与率から判断する限り、どの主題領域においても、登録されたすべての論文・記事が学術雑誌に掲載されるわけではない。付与率は全体で 5 割程度、領域別では物理学の理論系の領域が高いが数学や計算機科学では低く、領域間の差が大きい。
- 3) 全体として、*Physical Review* 各誌や *Astro-physical Journal* など、物理学と天文学のコアジャーナルに掲載された論文が多数登録されており、雑誌が扱う領域としては高エネルギー物理学、物性物理学、宇宙物理学関連の論文が多い。一部の学術雑誌ではほぼすべての掲載論文が arXiv に登録されている場合がある。arXiv の各主題領域には、その領域に対応した学術雑誌の掲載論文が多く登録されている。
- 4) 高エネルギー物理学 4 領域では早期アクセスを提供していると考えられるが、学術論文へのバックアクセスを目的とした登録も一部の主題領域で見られるようになっている。

仮説 1 の「arXiv に登録されたものは、最終的に何らかの成果公表メディアに掲載される」については、journal-ref 付与率から判断すると、付与率は最高でも 77.7% であることから、必ずしも arXiv に登録されるものすべてが学術雑誌をはじめとした何らかの成果公表メディアに掲載されているわけではないと考えられる。arXiv は公式には出版できない資料のためのリポジトリではないとされているが⁷⁴⁾、SLAC 図書館による journal-ref の修正が行われている高エネルギー物理学においても付与率が 100% ではないことを考えれば、登録された論文・記事が arXiv 以外には掲載されていないものもあると推測される⁷⁵⁾。こうした journal-ref が付与されていない論文・記事の実態およびそれらが利用者である研究者にどのように利用され認知されているかについては、今後検討する必要がある。

仮説 2 の「(journal-ref の付与された論文・記事の) その掲載先は、主に学術雑誌である」につ

いては、第 7 表の登録された論文・記事の掲載先から判断すると、arXiv 全体としては主として学術雑誌に掲載された論文が登録され、それ以外の会議録論文や書籍の一部は学術雑誌論文ほどには登録されていないと考えられる。ただし、主題領域別に見ると、掲載数上位 10 タイトルがすべて学術雑誌で占められる領域もあれば、計算機科学のように会議録やモノグラフシリーズが上位に現れる領域もある。journal-ref を付与されていても、必ずしも学術雑誌に掲載されているわけではなく、その中には会議録論文、講義録、博士・修士論文、書籍の一部などの出版物も含まれている⁷⁶⁾。journal-ref が付与されていないか、学術雑誌などに掲載されていないか、掲載されたが journal-ref を更新していないかのどちらかの場合が考えられるが、これらを考慮しても、現状では arXiv は純粋に学術雑誌投稿論文の原稿の電子版、すなわち e-print のアーカイブであるとは言いがたい。依然として学術雑誌投稿論文が中心であることに変わりはないが、より正確には掲載先が学術雑誌に限定されない学術論文一般の電子アーカイブになっていると考えられる。以下で述べるように、プレプリントが時間が経つことでそのまま古い論文として蓄積されるという意味での電子アーカイブではなく、登録時点ですでに刊行された論文を登録する電子アーカイブとしても機能している。

仮説 3 である「arXiv に登録されたものは、プレプリントであり学術雑誌に掲載される以前からアクセス可能である。既に学術雑誌に掲載済みの論文が登録されることはない」については、arXiv 全体を見れば、創設当初から一貫して、論文の掲載先が正式に刊行する前にアクセスを提供する機能を果たしていると考えられる。特に、高エネルギー物理学の領域においてはその傾向が顕著である。同領域では、他の領域と比較して掲載年以降に論文を登録することがほとんど見られない。しかし、計算機科学や数学といった主題領域では、登録以前にすでに掲載された過去の論文が一定数以上登録され、特に計算機科学では掲載年・前年に登録された論文の割合が 1 割程度と全体

で最も低く、これまで多くの文献で言及されてきた arXiv の代表的性質である即時性、すなわち提供される論文が掲載前のものであることを十分に満たしているとは考えられない。非線形科学、数量的生物学、物理学一般といった領域でも過去の論文を登録する割合が高いという点で異なる傾向にある。この結果から、高エネルギー物理学のように早期アクセスを提供する主題領域と計算機科学など過去の論文の登録率が比較的高い主題領域では、arXiv という同じプラットフォームを使っているものの、利用者間で登録される論文や登録行為そのものが異なる意味合いを持っていると考えられる。たとえば、計算機科学分野では、研究成果の迅速な流通が必要なため、成果公表メディアとしてテクニカルレポートや会議録論文が重要視される。研究者個人のウェブページや学部単位で設置されたリポジトリに学術雑誌論文やテクニカルレポートを電子的に登録し公開する慣習ができており、ACM (Association for Computing Machinery) や IEEE (電気電子学会) などは出版者版のファイルをウェブページやリポジトリで公開することを許可していた。しかし、それらに登録および公開された研究成果が保存される保証もなく、さらには成果を登録し検索できる単一のアーカイブが存在しなかった。1998 年 9 月に、ACM のイニシアチブにより、当時すでに設置されていた arXiv の計算機科学 (cs) のアーカイブと各大学に設置されたテクニカルレポートのリポジトリの横断検索システムである NCSTRL (Networked Computer Science Technical Reference Library: アンセストラル) とを組み合わせる形で、新たに CoRR (Computing Research Repository) が創設され、現在に至るまで arXiv の 1 アーカイブとして設置されている^{77)~80)}。以上のような分野の特性とアーカイブの設置経緯とが、計算機科学のアーカイブに登録された論文・記事の掲載先がその他の分野と異なる傾向にある一因だと考えられる⁸¹⁾。

論文への早期アクセスを提供している高エネルギー物理学における arXiv の登録実態がある意味本来の姿であるとすれば、過去の論文を登録す

る割合が高い主題領域では、arXiv はプレプリントに限定されず、刊行時期に関係なく論文を登録できる電子アーカイブとしても、一部の研究者によって利用されていると解釈できる。創設経緯を見れば、arXiv はあくまでプレプリントのサーバであり、こうした本来の目的とは必ずしも合致しない利用形態が生じている。arXiv が扱う領域が拡大し 20 年弱の時間を経ることで、arXiv とは何をいつ登録すべき場所なのかについての理解が研究者の間でも変化し一様ではなくなってきていると考えられる。以上から、仮説 3 は arXiv 全体および高エネルギー物理学をみれば該当するであろうが、必ずしも当てはまらない領域が出てきていると言える。

以上から、arXiv という学術情報メディアは、登録された論文・記事という点から見ると、“ネット上で、プレプリント (学術誌に掲載される前の論文) を収集・公開するためのサーバー”⁸⁾といったような一様な学術情報メディアでは必ずしもなく、二つの異なる様態を同時に併せ持ちつつあると考えられる。一つは、学術雑誌投稿論文の原稿を刊行前に電子的に登録・流通させるプレプリントサーバという創設当初からのものである。最も典型であるのは高エネルギー物理学分野で、一貫してその機能を果たしている。もう一つは、学術雑誌掲載論文だけに限らず、会議録論文や講義録など広く学術論文一般を刊行時期にかかわらず電子的に登録・流通させる電子アーカイブであり、計算機科学や数学の実態がこれに該当する。すなわち、主題領域が物理学系かそれ以外か、理論系かそれ以外かによって、arXiv の学術情報メディアとしての位置づけは異なり、各分野や研究領域における学術雑誌をはじめとした成果公表メディアの位置づけの違いがその背景にあると考えられる。SLAC などの連携機関がある高エネルギー物理学分野とそうした機関がない領域とでは、arXiv で論文・記事を流通させるための支援体制が異なっており、それが領域間の差につながっていると推測される。

ただし、本調査結果にはいくつか制限がある。journal-ref が付与された約 21 万件の登録論文・

記事を調査対象としており、付与率が低い主題領域においては本調査の結果と実態との間に誤差がある可能性がある。また、調査対象が arXiv に登録された論文・記事のみであり、今回見られた傾向や変化の原因についてまでは十分に明らかにすることができなかった。今後は、journal-ref が付与されていないものの実態調査や被引用調査、エンバーゴ方針など学術雑誌側からの調査、計算機科学、数学研究者などの利用行動や認識の調査など、より狭い主題領域ごとに調査を行う必要があるだろう。

謝 辞

査読者の方々からは、本稿の内容を改善する貴重なご指摘・ご意見を多数頂きました。本研究を進めるに当たって、データ処理で亜細亜大学国際関係学部のア形輝准教授から多くのサポートを受けました。論文の構想・執筆に当たっては、慶應義塾大学文学部の倉田敬子教授から多くの貴重なご指導・ご意見を頂きました。あわせてお礼申し上げます。本稿は、2008 年日本図書館情報学会春季研究集会（東京大学）における発表「学術情報メディアとしての arXiv の位置づけ」に基づいたものである。

注・引用文献

- 1) arXiv.org e-Print archive. <http://arxiv.org/>, (accessed 2008-10-17).
- 2) Stix, G. 電子アーカイブで科学界を結末. 日経サイエンス. 2003, vol. 7, p. 92-93.
- 3) 小山内正明. 電子ジャーナルと学術出版の未来. 情報の科学と技術. 1996, vol. 46, no. 7, p. 390-396.
- 4) Day, M. The scholarly journal in transition and the PubMed Central proposal. Ariadne. 1999, no. 21, <http://www.ariadne.ac.uk/issue21/pubmed/>, (accessed 2008-10-17).
- 5) Kling, R.; Spector, L. B.; Fortuna, J. The real stakes of virtual publishing: The transformation of E-Biomed into PubMed central. Journal of the American Society for Information Science and Technology. 2004, vol. 55, no. 2, p. 127-148.
- 6) Original Proposal for E-biomed (Draft and Addendum) E-BIOMED: A Proposal for Electronic Publications in the Biomedical Sciences. <http://www.nih.gov/about/director/pubmedcentral/ebiomedarch.htm>, (accessed 2008-10-17).
- 7) 科学技術・学術審議会 学術分科会 研究環境情報基盤部会・学術情報基盤作業部会. 学術情報基盤の今後のあり方について(報告) http://www.mext.go.jp/b_menu/shingi/gijyutu/gijyutu4/toushin/06041015/020.pdf, (入手 2008-10-17).
- 8) デイリー新語辞典. <http://dictionary.goo.ne.jp/search.php?MT=preprint+server>, (accessed 2008-10-17).
- 9) Taubes, G. Publication by electronic mail takes physics by storm. Science. 1993, vol. 259, p. 1246-1248.
- 10) Ginsparg, P. First steps toward electronic research communication. Computer in Physics. 1994, vol. 8, no. 4, p. 390-396.
- 11) Peskin, M. E. Reorganization of the APS Journals for the Era of Electronic Communication. <http://publish.aps.org/EPRINT/peskin.html>, (accessed 2008-10-17).
- 12) Matthews, R. Storming the barricades: For years, critics complained that peer review is a way of settling old scores and burying new research. Could the Internet smash the system?. New Scientist. 1995, no. 1982, p. 38.
- 13) Harnad, S. The post-Gutenberg galaxy: How to get there from here. Information Society. 1995, vol. 11, no. 4, p. 285-292.
- 14) Taubes, G. APS starts electronic preprint service. Science. 1996, vol. 273, no. 5237, p. 304.
- 15) Bolman, P. Journals face the electronic future. LOGOS. 1996, vol. 7, no. 1, p. 86-92.
- 16) Odlyzko, A. M. The slow evolution of electronic publishing. <http://www.dtc.umn.edu/~odlyzko/doc/slow.evolution.txt>, (accessed 2008-10-17).
- 17) 石山文彦. “高エネルギー物理学とインターネット(II)”. 物理屋のためのインターネット講座. パリティ編集委員会編. 丸善, 1997, p. 89-96.
- 18) 針谷喜久雄. “プレプリントサーバーに論文を投稿しよう”. 物理屋のためのインターネット講座. パリティ編集委員会編. 丸善, 1997, p. 61-70.
- 19) Luzi, D. E-print archives: A new communication pattern for grey literature. Interlending & Document Supply. 1998, vol. 26, no. 3, p. 130-139.
- 20) 高島寧. e-print archive から見る科学コミュニケーション・システム変容の可能性. 慶應義塾大学修士論文. 1999, 187 p.
- 21) 高島寧, 倉田敬子. “E-print Archive”. 電子メディアは研究を変えるのか. 倉田敬子編. 勁草書房, 2000, p. 139-172.

- 22) 倉田敬子, 松林麻実子. “物理学分野における動向”. 電子メディアは研究を変えるのか. 倉田敬子編. 勁草書房, 2000, p. 99-138.
- 23) Tomaiuolo, N. G.; Packer, J. G. Preprint servers: Pushing the envelope of electronic scholarly publishing. *Searcher*. 2000, vol. 8, no. 9, p. 53-61.
- 24) O'Connell, H. B. Physicists thriving with paperless publishing. *High Energy Physics Libraries Webzine*. 2002, no. 6, <http://library.cern.ch/HEPLW/6/papers/3/>, (accessed 2008-10-17).
- 25) Boyce, P. For better or for worse: Preprint servers are here to stay. *College & Research Libraries News*. 2000, vol. 61, no. 5, p. 405-407.
- 26) Peek, R. E-prints are gaining momentum: There are many tricky challenges in an Internet-speed world. *Information Today*. 2000, vol. 17, no. 9.
- 27) Quigley, B. Physics databases and the Los Alamos e-Print archive. *EContent*. 2000, vol. 23, no. 5, p. 22-26.
- 28) McKiernan, G. arXiv.org: the Los Alamos National Laboratory e-print server. *The International Journal on Grey Literature*. 2000, vol. 1, no. 3, p. 127-138.
- 29) Langer, J. Physicists in the new era of electronic publishing. *Physics Today*. 2000, vol. 53, no. 8, p. 35-38.
- 30) Till, J. E. Predecessors of preprint servers. *Learned Publishing*. 2001, vol. 14, no. 1, p. 7-13.
- 31) Brown, C. The coming of age of e-prints in the literature of physics. *Issues in Science and Technology Librarianship*. 2001, no. 31, <http://www.istl.org/01-summer/refereed.html>, (accessed 2008-10-17).
- 32) 倉田敬子. 学術情報流通の新たな方向性: 科学コミュニケーションと電子メディア. 情報知識学会誌. 2001, vol. 11, no. 1, p. 2-10.
- 33) Carrière, L. Jr. A brief history of e-prints and the opportunities they open for science librarian. *Science & Technology Libraries*. 2001, vol. 20, no. 2/3, p. 73-82.
- 34) Pinfield, S. How do physicists use an e-print archive? Implications for institutional e-print services. *D-Lib Magazine*. 2001, vol. 7, no. 12, <http://www.dlib.org/dlib/december01/pinfield/12pinfield.html>, (accessed 2008-10-17).
- 35) Garner, J.; Horwood, L.; Sullivan, S. The place of eprints in scholarly information delivery. *Online Information Review*. 2001, vol. 25, no. 4, p. 250-256.
- 36) Manuel, K. The place of e-prints in the publication patterns of physical scientists. *Science & Technology Libraries*. 2001, vol. 20, no. 1, p. 59-85.
- 37) Luce, R. E. E-prints intersect the digital library: Inside the Los Alamos arXiv. *Issues in Science and Technology Librarianship*. 2001, no. 29, <http://www.istl.org/01-winter/article3.html>, (accessed 2008-10-17).
- 38) Day, M. E-print services and long-term access to the record of scholarly and scientific research. *Ariadne*. 2001, no. 28, <http://www.ariadne.ac.uk/issue28/metadata/>, (accessed 2008-10-17).
- 39) Poynder, R. Continuing evolution in the world of scientific journal publishing. *Information World Review*. 2001, no. 166, p. 18-19.
- 40) Lawal, I. Scholarly communication: The use and non-use of e-print archives for the dissemination of scientific information. *Science & Technology Librarianship*. 2002, no. 36, <http://www.istl.org/02-fall/article3.html>, (accessed 2008-10-10).
- 41) Miranda, G. F.; Ginestet, J. The attitude of pharmaceutical industry research scientists to browsing and publishing on internet preprint and e-print servers. *Drug Information Journal*. 2002, vol. 36, no. 4, p. 831-837.
- 42) 九後太一. e-プリント・アーカイブの衝撃. 静脩. 2002, vol. 38, no. 3, p. 1-5.
- 43) 高田彰. 電子ジャーナル: 学術情報コミュニケーションの行方. 情報の科学と技術. 2002, vol. 52, no. 2, p. 83-87.
- 44) Harnad, S. “E-Prints: electronic preprints and postprints”. *Encyclopedia of Library and Information Science*. 2nd ed., Marcel Dekker, 2003, vol. 2, p. 990-993.
- 45) Butler, D. Biologists join physics preprint club. *nature*. 2003, vol. 425, p. 548.
- 46) Editorial. A place for preprint archives?. *nature neuroscience*. 2003, vol. 6, no. 5, p. 433.
- 47) O'Dell, C.; Dallman, D.; Vesely, M.; Vigen, J. Fifty years of experience in making grey literature available: Matching the expectations of the particle physics community. *Publishing Research Quarterly*. 2004, vol. 20, no. 1, p. 84-91.
- 48) Schwarz, G.; Kennicutt, R. C. Jr. Demographic and citation trends in astrophysical journal papers and preprints. *Bulletin of the American Astronomy Society*. 2004, vol. 36, no. 5, p. 1654-1663.
- 49) 倉田敬子. 科学技術情報流通の仕組み: 学術雑誌

- の役割. 情報の科学と技術. 2004, vol. 54, no. 6, p. 276-281.
- 50) Ginsparg, P. Scholarly communication architecture, 1989-2015. *Data Science Journal*. 2004, vol. 3, p. 29-37.
 - 51) Metcalfe, T. S. The rise and citation impact of astro-ph in major journals. *Bulletin of the American Astronomy Society*. 2005, vol. 37, no. 2, p. 555-557.
 - 52) Gunnarsdóttir, K. Scientific journal publications: On the role of electronic preprint exchange in the distribution of scientific literature. *Social Studies of Science*. 2005, vol. 35, no. 4, p. 549-579.
 - 53) 松林麻実子, 倉田敬子. e-print archive という情報メディア: 日本の物理学研究者への調査に基づいて. *日本図書館情報学会誌*. 2005, vol. 51, no. 3, p. 125-140.
 - 54) Hey, T.; Hey, J. e-Science and its implications for the library community. *Library Hi Tech*. 2006, vol. 24, no. 4, p. 515-528.
 - 55) Henneken, E. A.; Kurtz, M. J.; Eichhorn, G.; Accomazzi, A.; Grant, C.; Thompson, D.; Murray, S. S. Effect of e-printing on citation rates in astronomy and physics. *Journal of Electronic Publishing*. 2006, vol. 9, no. 2, <http://hdl.handle.net/2027/spo.3336451.0009.202>, (accessed 2008-10-17).
 - 56) Mele, S.; Dallman, D.; Vigen, J.; Yeomans, J. Quantitative analysis of the publishing landscape in high-energy physics. *Journal of High Energy Physics*. 2006, no. 12, SO1, doi: 10.1088/1126-6708/2006/12/S01, <http://www.iop.org/EJ/abstract/1126-6708/2006/12/S01>, (accessed 2009-03-31).
 - 57) Thomas, S. E. Publishing solutions for contemporary scholars: The library as innovator and partner. *Library Hi Tech*. 2006, vol. 24, no. 4, p. 563-573.
 - 58) Hanauske, M.; Bernius, S.; Dugall, B. Quantum game theory and open access publishing. *Physica A*. 2007, vol. 382, p. 650-664.
 - 59) 倉田敬子. “オープンアクセスとは何か”. 学術情報流通とオープンアクセス. 勁草書房, 2007, p. 145-186.
 - 60) Davis, P. M.; Fromerth, M. J. Does the arXiv lead to higher citations and reduced publisher downloads for mathematics articles?. *Scientometrics*. 2007, vol. 71, no. 2, p. 203-215.
 - 61) Ginsparg, P. Next-generation implications of open access. *CTWatch Quarterly*. 2007, vol. 3, no. 3, p. 11-18.
 - 62) ACS Journal Editor's Policy on Preprints. <http://pubs.acs.org/instruct/preprints.html>, (accessed 2008-10-17).
 - 63) Posting Articles on the Internet. <http://www.apa.org/journals/authors/posting.html>, (accessed 2008-10-17).
 - 64) 奥村晴彦. LATEX2ε 美文書作成入門. 技術評論社. 2007, 433 p.
 - 65) 森岡倫子. 電子ジャーナル黎明期の変遷: 1998年から2002年までの定点観測. *Library and Information Science*. 2005, no. 53, p. 19-36.
 - 66) Ginsparg, P. The global-village pioneers. *Physics World*. 2008, vol. 21, no. 10, <http://physicsworld.com/cws/archive/print/21/10>, (accessed 2008-10-17).
 - 67) arXiv:all Submissions '05-'07. http://arxiv.org/Stats/au_all.html, (accessed 2008-10-17).
 - 68) 石山文彦. “リサーチ・オンライン”. パリティ編集委員会編. 物理屋のためのインターネット講座. 丸善, 1997, p. 117-130.
 - 69) Butler, D. Los Alamos loses physics archive as preprint pioneer heads seat. *nature*. 2001, vol. 412, p. 3-4.
 - 70) arXiv.org help—To Withdraw a Paper. <http://arxiv.org/help/withdraw>, (accessed 2008-10-17).
 - 71) Birnholtz, J. P. What does it mean to be an author? The intersection of credit, contribution, and collaboration in science. *Journal of the American Society for Information Science and Technology*. 2006, vol. 57, no. 13, p. 1758-1770.
 - 72) Hurd, J. M. “High Energy Physics”. From Print to Electronic: The Transformation of Scientific Communication. Crawford, S. Y.; Hurd, J. M.; Weller, A. C. Medford, NJ, Information Today, 1996, p. 65-76.
 - 73) arXiv.org help—To Add a Journal-ref to a Paper. <http://arxiv.org/help/jref>, (accessed 2008-10-17).
 - 74) arXiv.org help—arXiv Primer. <http://arxiv.org/help/primer>, (accessed 2008-10-17).
 - 75) たとえば, hep-th/9808092 は, *Physics Letters B* に却下された論文である.
 - 76) ギンスバーグ自身でさえも, *New York Times* に寄稿した記事を登録している (physics/0408033).
 - 77) Halpen, J. Y. A computing research repository. *D-Lib Magazine*. 1998, vol. 4, no. 11, doi: cnri.dlib/november98-halpern, <http://www.dlib.org/dlib/november98/11halpern.html>, (accessed 2009-1-5).
 - 78) Halpen, J. Y.; Lagoze, C. “The computing research repository: Promoting the rapid dissemination and archiving of computer science

- research". Proceedings of the Fourth ACM Conference on Digital Libraries. 1999, p. 3-11.
- 79) Halpen, J. Y. CoRR: A computing research repository. ACM Journal of Computer Documentation. 2000, vol. 24, no. 2, p. 41-48.
- 80) Borgman, C. L. Scholarship in the Digital Age: Information, Infrastructure, and the Internet. Cambridge, Massachusetts, MIT Press, 2007, 336 p.
- 81) CoRR は計算機科学分野全体を対象にした中央型リポジトリとなることを目的としているものの、NSCTRL は現在静止状態である⁸⁰⁾。同分野の著名なデータベースである CiteSeerX が索引付けを行っている文書数約 120 万件と比較すると、2007 年末までに arXiv に登録された全論文・記事数約 9 千件は非常に少ないと言える。CoRR のこうした位置づけが、登録論文・記事の傾向に反映している可能性もある。

要 旨

【目的】本稿の目的は、プレプリントサーバ arXiv の学術情報メディアとしての役割および位置づけを、2007 年までに登録された論文・記事すべてに基づいて実証的に明らかにすることである。

【方法】1) 論文・記事の年間および総登録件数、2) 論文・記事の学術雑誌掲載率とその後の掲載先、3) arXiv への登録年と学術雑誌掲載年の差（即時性）を明らかにするために、1991 年から 2007 年にわたって arXiv に登録された論文・記事 449,071 件すべてを対象に、各論文・記事に付与された 1) ID, 2) journal-ref（登録論文・記事のその後の掲載先）、3) 主題領域のデータを分析した。

【結果】方法で述べた 3 点について、以下のことがわかった。1) 17 年間にわたって arXiv 全体では登録件数が毎年増加傾向にあり、とくに宇宙物理学、物性物理学、数学で顕著であるが、主題領域により差がある。2) arXiv に登録された論文・記事のその後の学術雑誌掲載率は全体で 47.1% であり、少なくとも論文・記事の半数弱だけが学術雑誌等に掲載されていること、領域では物理学の理論系の領域で高いが数学や計算機科学では低いこと、学術雑誌に掲載された登録論文・記事全体の半数近くを *Physical Review* 各誌や *Astronomical Journal* など、物理学と天文学のコアジャーナルに掲載された論文が占めていることがわかった。3) 高エネルギー物理学 4 領域では多くの論文が雑誌掲載年以前に登録されているが、登録時点で既に掲載済みである学術論文一般へのアクセス提供を目的としたものも一部の主題領域で見られるようになっている。以上から、現在、arXiv は学術雑誌論文のプレプリントサーバと学術論文一般の電子アーカイブという二つの機能を同時に果たしていると考えられる。