Citing e-prints on arXiv
A study of cited references in WoS-indexed journals from 1991-2013

Valeria Aman[1]

[1] *aman@forschungsinfo.de*
iFQ - Institute for Research Information and Quality Assurance, Schützenstraße 6a, 10117 Berlin (Germany)

Abstract

This study deals with the analysis of cited references in Web of Science (WoS) to e-prints on arXiv. Created in 1991, arXiv accelerated the scholarly communication and developed into a well-established e-print repository that functions as an essential access point to the latest research in physics, astrophysics, mathematics, computer science and related fields. Authors evidently rely on arXiv full texts and refer to them in their own research papers. These cited references to arXiv that represent the acceptance of e-prints in journals and series indexed in WoS are tackled in this paper. A total of 900,000 cited references to arXiv have been identified for the 1991-2013 period. Object of investigation is on the one hand the set of cited references to arXiv, and on the other hand the set of papers in WoS that cite arXiv. Among other things, the paper illustrates that citations to arXiv peak in the year after submission and drop rapidly. The geographical distribution of authorship citing arXiv in their papers shows that authors from the US, Germany, GB, France and Italy rely heavily on arXiv. The paper identifies "arXiv-friendly" journals where the majority of articles refer to arXiv.

Conference Topic
Journals, databases and electronic publications

Introduction

The arXiv is a convenient vehicle to disseminate research results prior to the publication of peer-reviewed articles. It is also common to submit postprints for reasons of wide availability and archiving. There is no doubt that e-prints are read by a wide community and are regarded to be of good quality. Thus, it is of interest to learn more about the perception of arXiv as a source of relevant information that supports researchers' ideas and discoveries. The study sets out to answer the following questions: 1) Do authors publishing in journals covered by Web of Science (WoS) cite e-prints on arXiv? 2) What characteristics in citations can be observed? 3) In which countries are authors situated that rely on e-prints in arXiv? 4) What are the journals that include the highest rate of articles with cited references to arXiv?

Background

The rise of preprints, e-prints and arXiv

There are several definitions for the term "preprint". Lim (1996) defines a "preprint" as a manuscript that has been reviewed and accepted for publication, a manuscript that has been submitted for publication, but for which a decision to publish has not been made yet, or a manuscript that is intended for publication, but is being circulated for comments among peers prior to journal submission. Electronic prints (e-prints) refer both to preprints and post-prints (peer-reviewed published papers), and other documents that are made available on the Internet. The "preprint culture" dates back to the 1960ies, when high-energy physicists were eager to disseminate their results by printing and mailing copies of their manuscripts simultaneously to journal submission (Goldschmidt-Clermont, 1965). The time consuming process of peer-review was hence effectively bypassed. With the advent of the World Wide Web in the early 1990ies, the emergence of new methods of scientific discourse were encouraged, altering the traditional channels of scholarly communication (Brown, 2001).

In summer 1991, Paul Ginsparg conceived the repository arXiv at the Los Alamos National Laboratoy (LANL) in New Mexico. Ginsparg (1994, p.157) stated that "the realization of arXiv was facilitated by a pre-existing 'preprint culture', in which the irrelevance of refereed journals to ongoing research has long been recognized". Ginsparg (1994, p.159) designed arXiv (formerly xxx.lanl.org) as a fully automated system, where users could maintain a database to disseminate information without outside intervention.

Originally, arXiv was intended for the High-Energy Physics (HEP) community, but expanded rapidly to cover all of Physics, Astrophysics, Mathematics and Computer Science. Since September 2003 arXiv covers Quantitative Biology. In April 2007 Statistics was included, followed by Quantitative Finance in December 2008. Today, arXiv is hosted at Cornell University in New York with seven mirror sites all over the world. It contains more than 1,000,000 full-text e-prints, receiving about 9,000 new submissions each month.[1] Researchers can check arXiv for new information, search for relevant papers, post their own papers and cite references by arXiv ID. It is a self-organizing publication mode that costs the users nothing (Langer, 2000). Another reason for arXiv's popularity is its democracy, because scientists "can post their research results without being hassled by grumpy editors and referees" (ibid., p.35). According to Ginsparg (1994, p.157) physicists have learned to determine from the author, title and abstract whether to read a paper "rather than rely on the alleged verification of overworked or otherwise careless referees".

Nowadays, researchers still regard it as valuable to publish their work in peer-reviewed journals. Prior to formal publication, the findings may be spread as conference proceedings, reports, working papers or preprints. As Heuer, Holtkamp and Mele (2008, p.2) point out "scientists expect unrestricted access to comprehensive scientific information in their field, state-of-the-art information venues to optimize their research workflow and quality assurance at the parallel existence of traditional peer-review and the immediacy of dissemination and feedback". A publication delay of several months between the completion of a work and its appearance in a peer-reviewed journal is simply a "negative phenomenon in scientific information dissemination" (Amat, 2008, p.379). Amat (ibid.) found that the publication delay depends primarily on the peer-review process (see also Luwel, 1998). ArXiv serves to overcome this delay and helps to circulate results upon realization.

Previous work

The citation behaviour of e-prints available through arXiv has been studied extensively. Youngen (1998) identified the growing importance of e-prints in the published literature. He found that e-prints became the first choice among physicists and astronomers for finding current research and keeping up with colleagues and competitors at other institutions. Brown (2001) studied citations of e-prints on arXiv in astronomy and physics journals from 1998 to 1999. The citation analysis showed that the peak of citations to e-prints is reached after three years, which is comparable to papers in print journals. Garner, Horwood & Sullivan (2001) determined the place of e-prints in the scholarly information delivery, concluding that rapid dissemination of results in form of preprints establishes priority and enables rapid feedback. Brown (2003) asked for the opinion of chemists about citing e-prints in the articles they author. Fifty-two percent said they would cite e-prints whenever possible, whereas 48% stated that they would not. Reasons for avoiding to cite the Chemistry Preprint Server (CPS) are the lack of relevant articles, the lack of customary to cite, and the lacking awareness of CPS (ibid., p.365). The study of infiltration of CPS e-prints into the literature of chemistry revealed that "no citations to e-prints were found in the journal literature using ISI's Web of Science from 2000 to 2001" (ibid., p.366). Prakasan & Kalyane (2004) focused on the

[1] http://arxiv.org/stats/monthly_submissions / [Last visited January 06, 2015]

citations in Science Citation Index to e-prints on arXiv, submitted under the four categories hep-ex, hep-lat, hep-ph and hep-th[2], providing a broad insight into citation habits.

Several studies focused on the citation impact of e-prints on arXiv, also within the Open Access debate (see Harnad & Brody, 2004; Antelman, 2004). Schwarz & Kennicutt (2004) analyzed articles published in the Astrophysical Journal in 1999 and 2002 and reported that papers posted to the astro-ph-section on arXiv were cited more than twice as often as those without a version on arXiv. In accordance, Metcalfes (2005) findings show that astronomy papers in the highly-cited journals Science and Nature received higher citation rates when their authors posted their papers on arXiv's astro-ph. Metcalfe (2006) studied the field of solar physics with the result that papers posted to arXiv are on average 2.6 times as often cited as papers not being posted. He concludes that higher citation rates are not a result of self-selection of outstanding papers, since conference proceedings reveal the same result. Moed (2007) analyzed how the citation impact of articles deposited in the Condensed Matter section in arXiv and subsequently published in a journal compares to that of articles not deposited on arXiv. He concluded that arXiv accelerates citations, because it makes papers earlier available. Davis & Fromerth (2007) examined whether mathematics journals from 1997 to 2005 with a previous preprint version on arXiv receive more citations than non-deposited. Their findings show that articles in arXiv receive on average 35% more citations, which translates to 1.1 citations per article. They explain the citation advantage with the Open Access, the Early View, and the Quality postulates, which are non-exclusive.

Henneken et al. (2007) analyzed whether e-prints on arXiv are preferred over the journal articles in four core journals in astrophysics. They found that as soon as an article is published, the community prefers to read and cite it, so that the usage in the NASA Astrophysics Data System (e-print system) drops to zero. They also showed that the half-life (the time at which the use of an article is half the use of a newly published article) for an e-print is shorter than for a journal article. Gentil-Beccot, Mele & Brooks (2009) investigate whether HEP scientists still read journals or rather prefer digital repositories. Their citation analysis shows that free and immediate dissemination of preprints results in a citation advantage for HEP journals. Furthermore, their analysis of clickstreams reveals that high-energy physicists prefer preprints and seldom read journals.

Some of the studies suggest that articles with a previous preprint on arXiv receive more citations than articles without. Other studies report no such effect. Gentil-Beccot, Mele & Brooks (2009) did not detect any citation advantage from publishing in Open Access HEP journals. Their finding is similar to that of Moed (2007) in Condensed Matter, Davis (2007) in Mathematics and Kurtz & Henneken in Astrophysics (2007).

Brody, Harnad & Carr (2006) examined the correlation of the number of article downloads and the number of citations. On the basis of arXiv they show that the short-term Web usage impact of e-prints predicts a medium-term citation impact of the final article. Haque and Ginsparg (2009; 2010) found that e-prints posted to arXiv at the beginning and end of a day reach a wider readership and receive higher citation rates over the course of ensuing years than posting in the middle of day. Shuai, Pepe & Bollen (2012) analyzed the online response to preprint publications on arXiv, studying the delay of article downloads and Twitter mentions following submission.

Larivière et al. (2014) analyzed the proportion of papers across all disciplines on arXiv for the 1991-2012 period, just as the proportion of arXiv papers that are published in WoS-indexed journals. They determine the time between arXiv submission and journal publication, ageing characteristics and impact of arXiv e-prints and their published alter ego. They also focus on

[2] High energy physics - experiment (hep-ex), high energy physics - lattice (hep-lat), high energy physics - phenomenology (hep-ph), and high energy physics - theory (hep-th).

the proportion of cited references in WoS to arXiv e-prints by discipline. Working with percentages, they quantify that journals in nuclear and particle physics have 6.6% of their references to arXiv e-prints, whereas in mathematics this share is below 1.5% (ibid., p.1163). Stimulated by the work of Larivière et al. (2014), this study sets out to quantify the number of cited references in WoS to arXiv manuscripts, and to provide a broader view on characteristics of cited references and the papers that include them.

Data and methods

Database

The study builds upon the bibliometric database at the "Competence Center for Bibliometrics for the German Science System" that is hosted at the iFQ.[3] It consists of data from Thomson Reuter's Web of Science. Peer-reviewed journal articles are the primary mode of communication of scientific research. Researchers write reviews or articles with discoveries, theories and results. To relate their work they cite other articles if they know the article and believe it to be relevant to their own work. They might also provide negative citations in order to disagree or to say that a paper has flaws (see Brody, Harnad & Carr, 2006). Citations can be therefore used as a measure of influence and importance of preceding articles.

The identification of references to arXiv depends on the quality of the bibliographic information (e.g. the presence of the reference to arXiv) and the extent to which WoS was able to parse the references of the citing articles. Identifying cited references to arXiv can lead to false positives, when a reference looks like an arXiv identifier but is actually not, or where authors make mistakes. A linking by bibliographic data is more precise as it builds upon author names, journal title, volume, page number, year of publication etc.

Data collection

Different from Youngen (1998), who analyzed those cited references that state explicitly "preprint" in ISI's SciSearch (p.451), this study also includes postprints. Hence, all manuscripts on arXiv are in the following referred to as "e-prints". The e-print identifier assigned by arXiv provides a standardized number that allows each e-print to be uniquely identified. This uniqueness is required for correct citing of the work. ArXiv has established a subject grouping and numbering system for submitted e-prints. Examples are Astrophysics (astro-ph), Condensed Matter (cond-mat), High-Energy Physics-Theory (hep-th) or Nuclear-Experiment (nucl-ex), followed by a numerical string, indicating the year and month of submission, and an increasing accession number. A typical example is quant-ph/95002, where quant-ph stands for Quantum Physics, "95" for the year 1995 and "002" for the accession number. Up to March 2007 this ID enabled a broad subject categorization. In April 2007, the arXiv-ID was changed and no longer contains subject categories. It consists of eight digits, of which the first four represent the year and month of submission. Divided by a period, they are followed by a four-digit long accession number, e.g.: arXiv: 0705.0002. We can infer that this e-print was loaded in May 2007. Since the accession number will soon reach its capacity, the length of the accession number has been extended by one digit in January 2015.[4]

The search for arXiv e-prints in the cited reference field in WoS was approached in several steps. E-prints up to 2007 were identified on the basis of an alphanumeric string that contains the subject category followed by the year of submission and the accession number.[5] E-prints published in 2007 or later were identified by the string "arXiv" followed by a numerical string. This led to an overall satisfying result, since the string "arXiv" is unique and causes

[3] http://www.bibliometrie.info/ [Last visited January 06, 2015]
[4] http://arxiv.org/new#dec19_2014 [Last visited January 06, 2015]
[5] The categories in bold print were used for the matching: http://arxiv.org/ [Last visited January 06, 2015]

almost no confusion. A low number of false positives cited references were deleted manually. Only one in four cited references had a publication year assigned, which is indeed not necessary, since it is part of the arXiv ID. With the application of Regular Expressions in SQL the year of e-print publication was deduced for more than 99% of cited references. A publication year was not deducible, where authors cited arXiv simply in this fashion: "arXiv". The search strategy may not include citations to works that technically have to be considered as arXiv e-prints. According to Youngen (1998, p.451) authors may have cited preprints as "submitted to...", "to be published in...", "in press" or "unpublished", depending on their state in the publication cycle. Thus, in reality, the number of citations to e-prints on arXiv may be much higher than presented here.

Data corpus[6]

With the search strategy described, 892,867 cited references to arXiv were identified for the 1991-2013 period, of which 357,557 have a distinct character string. Due to multiple subject categorizations in arXiv, author typos, or erroneous data parsing in WoS, one and the same e-print can be referred to in different spelling variants. Hence, the actual number of arXiv e-prints cited in the 1991-2013 period by papers in WoS is lower. At the same time 289,145 distinct papers were identified in WoS that constitute these 892,867 cited references. To relate these figures, Brown (2001) found 35,928 citations to arXiv e-prints (posted between 1991 and 1999) in astronomy and physics journals published in 1998-1999. In the following, analyses are based on the cited references to arXiv and the WoS-papers that include them.

Results and discussion

Figure 1 provides an overview of the data collected. The number of e-prints submitted to arXiv has been gradually rising from 303 in 1991 to 92,641 in 2013.[7] The number of papers in WoS citing at least one e-print on arXiv has steadily increased and comprises around 28,000 papers in 2013. In addition, we can see the number of cited references to e-prints on arXiv with the publication year of the citing paper as indicated on the x-axis. We can derive that a paper citing arXiv includes on average more than one citation to e-prints on arXiv. Most of the citations to e-prints were provided in 2012 (ca. 76,000).

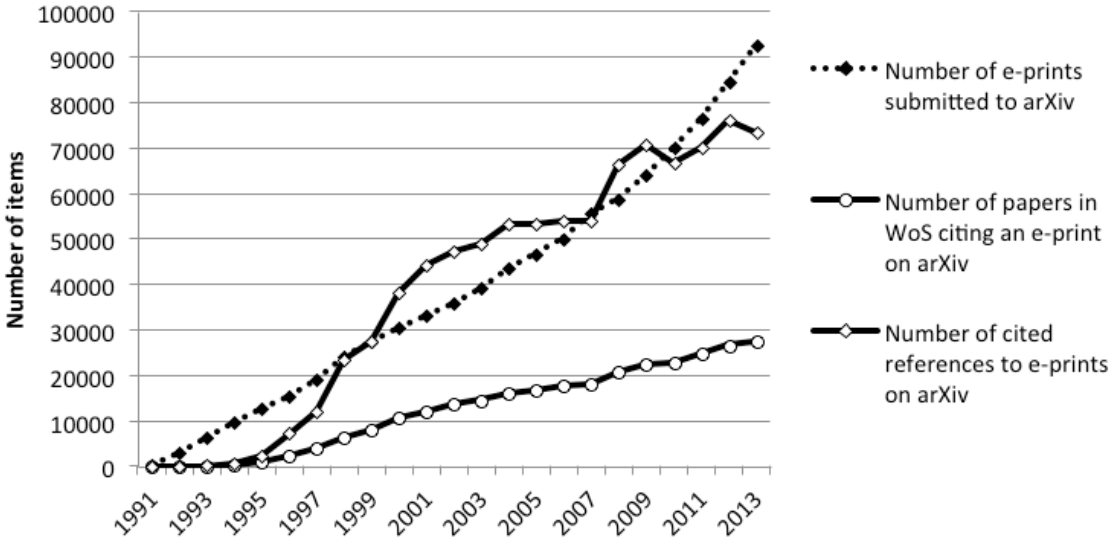

Figure 1. Overview of the yearly growth of submissions to arXiv, the number of papers in WoS citing arXiv e-prints according to their publication year, and the number of cited references.

[6] The data corpus can be requested on demand.
[7] http://arxiv.org/stats/monthly_submissions [Last visited January 06, 2015]

The analysis of document types shows that articles rank first with 96.0% of all WoS documents from 1991-2013 that cite arXiv. Reviews (3.2%) refer to arXiv as well, in order to provide a broad or up-to-date state of research. Editorials, Letters, Corrections and Notes also reference arXiv.

In the following, it does make a difference whether cited references are analysed or the WoS-papers that include those. Due to different citation habits, even within a broad field such as physics, it appears more suitable to consider primarily the citing papers. Table 1 provides an overview of the subject areas that constitute most of the citations to arXiv. The first column lists the Subject Categories[8] (SC) in WoS in a descendant order, regarding the number of arXiv citing papers assigned to this SC. We can see that Particle Physics ranks first (21%), followed by Astronomy and Astrophysics. In total, these 12 SC cover more than 90% of all citing papers that refer to arXiv between 1991 and 2013. The percentages and order of the SC changes when we have a look on the number of cited references to arXiv. Particle Physics still ranks first, claiming almost one-third of all cited references to arXiv. The results suggests that papers in Particle Physics have on average a higher number of cited references to arXiv than those in other SC.

Table 1. Overview of Subject Categories in WoS that contribute to the majority of papers that cite arXiv and their number of cited references. The data is based on 289,145 arXiv-citing papers in WoS that provide 892,867 cited references in 1991-2013.

Subject Category	No. of papers citing arXiv	Share in %	No. of cited references	Share in %
Physics, Particles & Fields	88,757	21.0	398,022	30.5
Physics, Multidisciplinary	70,383	16.7	248,091	19.0
Astronomy & Astrophysics	68,805	16.3	225,326	17.3
Physics, Mathematical	28,073	6.7	82,490	6.3
Physics, Condensed Matter	25,658	6.1	49,852	3.8
Mathematics	23,894	5.7	46,952	3.6
Physics, Nuclear	22,838	5.4	83,712	6.4
Optics	13,602	3.2	27,414	2.1
Physics, Atomic, Molecular & Chemical	12,754	3.0	25,625	2.0
Mathematics, Applied	10,976	2.6	20,169	1.5
Physics, Applied	9,223	2.2	17,099	1.3
Physics, Fluids & Plasmas	5,704	1.4	9,488	0.7

This leads us to the analysis of the distribution of cited references among the papers in WoS that cite arXiv. Table 2 illustrates the frequency of citing papers in WoS that include as many cited references as stated in the left column. We can see that six papers in WoS have more than 200 references to arXiv in their list of references. Every eleventh paper, out of the set of arXiv citing papers, includes 6 to 10 references to arXiv. Nevertheless, around 46% of citing papers provide a single reference to arXiv. A closer look on the paper with the highest number of cited references to arXiv shows that it is a review article from 2000 on String Theory and Gravity, where a link to arXiv was set additionally to the journal article reference. This brings us to the analysis of characteristics in citations to arXiv. Are e-prints on arXiv immediately cited when there is no corresponding journal article or are they also used in future and even preferred over the corresponding journal article?

[8] The 260 SC in WoS are assigned to journals on the basis of their scope and citation links.

Table 2. Distribution of cited references among WoS-papers that cite e-prints on arXiv.

Number of references to arXiv in a single paper	Number of papers citing arXiv	%
more than 200	6	0.00
151 to 200	8	0.00
101 to 150	29	0.01
51 to 100	222	0.08
21 to 50	2,567	0.89
11 to 20	9,375	3.24
6 to 10	25,859	8.94
5	12,544	4.34
4	18,939	6.55
3	30,969	10.71
2	56,204	19.44
1	132,423	45.80
Total	289,145	100.00

Figure 2 shows on the one hand the line graph of all citations to e-prints on arXiv up to 2013. Different from Figure 1 the x-axis signifies the year of e-print publication. Thus, the sudden decrease of cited e-prints from 2008 on is due to the fact that they had less time to be referenced than those posted in earlier years. In addition, Figure 2 provides bars indicating the years in which these e-prints were cited by WoS papers. Each bar represents the number of cited references to arXiv in the same year as the e-print was published, the subsequent year and two and three years respectively after publication of the e-print. The space between the line graph and the bars represents the cited references to e-prints that were provided more than three years after e-print publication. Since e-prints from recent years did not have much time to be cited, the bars coincide with the line graph of the total number of cited e-prints.

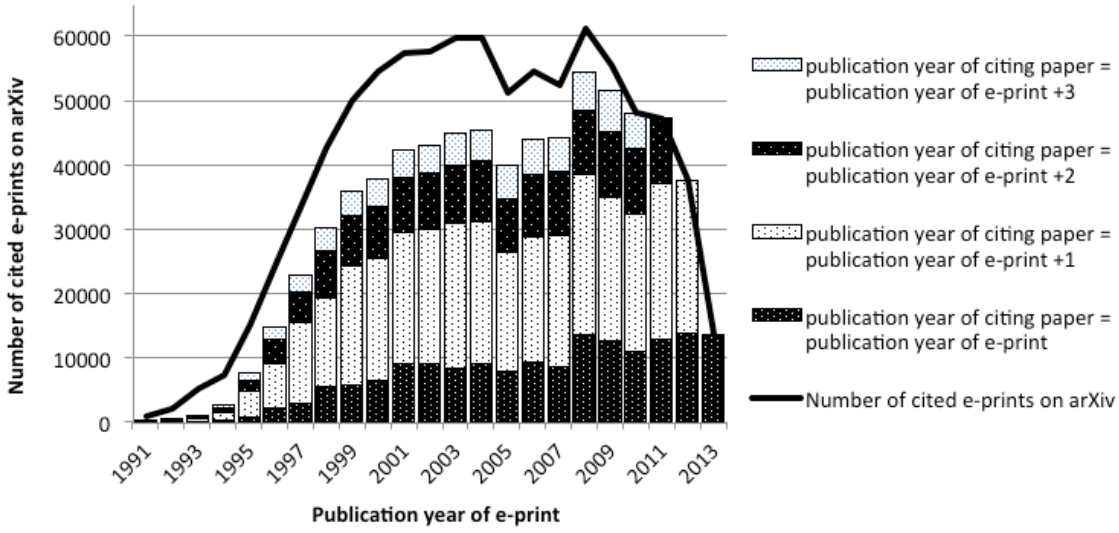

Figure 2. Time series of citation distribution. Illustrated are citations that equal the year of e-print submission, citations to e-prints that are one year old, up to the age of three years. The line graph signifies the total number of e-prints cited, published in the year as indicated.

It becomes evident that e-prints on arXiv are mostly cited in the subsequent year of e-print post. Almost half of all cited references in a year relate to e-prints that were placed on arXiv the preceding year. This is in accordance with Larivière et al. (2014, p.1166), who found that citations to e-prints on arXiv peak the year following submission. The figure also indicates that e-prints are cited immediately in the same year of posting. Only a small share of cited

references points to three-year old e-prints. On the contrary, Brown's (2011) analysis in astronomy and physics showed that the peak of citations to e-prints is reached after three years. The results in Figure 2 are in little accordance with Henneken et al. (2007, p.19) who showed that the usage of e-prints drops to zero as soon as the journal article has appeared, suggesting that authors have access to subscribed journals and prefer to cite the refereed version. Garner, Horwood & Sullivan (2001, p.251) quantified that 90% of papers on arXiv are later published in journals so that a corresponding article can be found and cited properly.

Nevertheless, there are many reasons that underscore the high citation rates of e-prints. Davis & Fromerth (2007) write that the arXiv copy is sufficient for the purpose of citing it in one's own work. They found that articles that are also accessible on arXiv receive 23% fewer downloads from the publisher's web site two years after publication (ibid., p.23). Gentil-Beccot, Mele & Brooks (2009) found that citations start before publication, because scientists in HEP do not wait for an article to be published. Even in the first few months after journal publication authors read and cite the preprint (ibid., p.6). According to Moed (2007) colleagues start to read a paper and cite it in their own articles earlier if it is deposited on arXiv. The following Figure 3 illustrates the relation between the publication year of a WoS-paper citing arXiv, and the publication year of the cited e-print. The whole bar in each year (y-axis) represents the total number of cited references to e-prints on arXiv from this year (cf. Figure 1). The cited references from each year are grouped by the publication year of the cited e-print. Each bar indicates the share of e-prints, according to their year of publication. For the year 2013 we can see that 13,000 cited references (top black part of the 2013-bar) refer to e-prints published in the same year. The lion's share of cited references in 2013 (24,000) is to e-prints published in 2012. In general, we can conclude from Figure 3 that the majority of references in each year points to e-prints published in the preceding year.

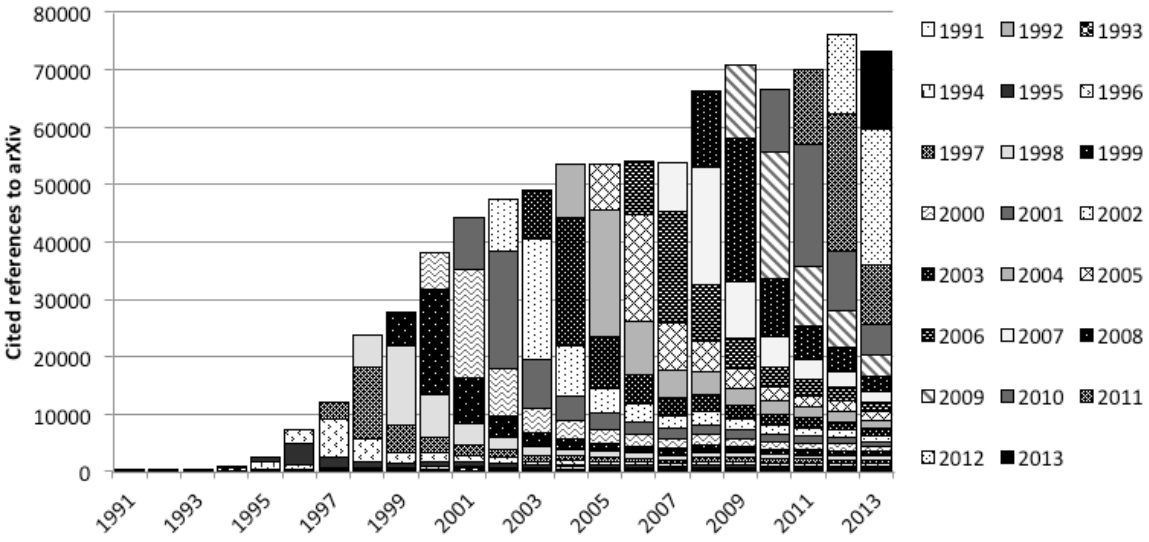

Figure 3. Time series of cited references to e-prints on arXiv. The x-axis represents the publication years of WoS-paper citing an e-print, whereas each bar represents the share of the years a cited e-print was published in.

To see where the authors that frequently cite arXiv are from, Table **3** provides a ranking of countries according to the highest number of papers in WoS with at least one cited reference to arXiv. USA rank first with one-third of all papers that cite arXiv. They are followed by Germany and Great Britain. Note that the percentages do not add up to 100, since co-authored papers can be attributed to multiple countries.

Table 3. Overview of countries that most frequently cite arXiv e-prints. The percentages are calculated on the basis of the total number of citing papers (289,145).

Rank	Country	No. of WoS-papers citing e-prints	%	Rank	Country	No. of WoS-papers citing e-prints	%
1	USA	97,085	33.6	11	Switzerland	14,489	5.0
2	Germany	45,842	15.9	12	India	11,764	4.1
3	GB	30,776	10.6	13	Poland	9,332	3.2
4	France	28,159	9.7	14	Brazil	9,004	3.1
5	Italy	27,896	9.6	15	Netherlands	8,361	2.9
6	China	25,467	8.8	16	South Korea	8,271	2.8
7	Japan	25,196	8.7	17	Australia	7,296	2.5
8	Russia	22,772	7.9	18	Israel	7,019	2.4
9	Spain	15,902	5.5	19	Sweden	5,402	1.9
10	Canada	14,879	5.1	20	Belgium	4,709	1.6

The journals whose articles most often cite e-prints on arXiv are identified in Table 4. On the left of the table, journals are ranked according to their number of citing papers in the 1991-2013 period. On the right of the table journals are ranked according to their number of cited references to arXiv. Evidently, most of the journals carry a majority of HEP content. Among these are *Physical Review D*, *Journal of High Energy Physics* (JHEP), *Physics Letters B* and *Nuclear Physics B*. Striking are also the astrophysical journals, among which we can find the *Astrophysical Journal, Monthly Notices of the Royal Astronomical Society* and *Journal of Cosmology and Astrophysical Physics*.

Table 4. Overview of journals in WoS with the highest number of papers citing arXiv and journals with most of the cited references to arXiv in the 1991-2013 period.

Journal	Citing papers	%	Journal	Cited ref.	%
Physical Review D	30,287	10.5	Physical Review D	112,261	12.6
Physical Review B	15,080	5.2	Journal of High Energy Physics	77,431	8.7
Journal of High Energy Physics	14,881	5.1	Physical Review B	66,750	7.5
Physical Review Letters	13,816	4.8	Nuclear Physics B	50,757	5.7
Physics Letters B	13,707	4.7	Physics Letters B	29,195	3.3
Physical Review A	9,599	3.3	Physical Review Letters	28,873	3.2
Astrophysical Journal	8,428	2.9	Classical and Quantum Gravity	22,969	2.6
Nuclear Physics B	8,033	2.8	Physical Review A	20,480	2.3
Monthly Notices of the Royal Astronomical Society	6,256	2.2	Journal of Cosmology and Astrophysical Physics	19,559	2.2
Physical Review E	5,081	1.8	International Journal of Modern Physics A	18,685	2.1
Sum	125,168	43.3	Sum	446,960	50.1

Youngen (1998) could not find firm rules for citing preprints, with the exception of the Astrophysical Journal, which stated that "References to private communications, papers in preparation, preprints, or other sources generally not available to readers should be avoided" (p.453). Nevertheless, it ranks seventh among the most active journals citing e-prints on arXiv. This restriction must have been eased over the years, as can be seen in Figure 4. Depicted are time series of percentages of papers in a journal that cite arXiv, for the ten

journals with the highest number of arXiv-citing papers (see Table 4). We can observe that up to 1997 the *Astrophysical Journal* had less than 10% of their papers citing e-prints on arXiv. This share was growing in the following years to reach approx. 25%.

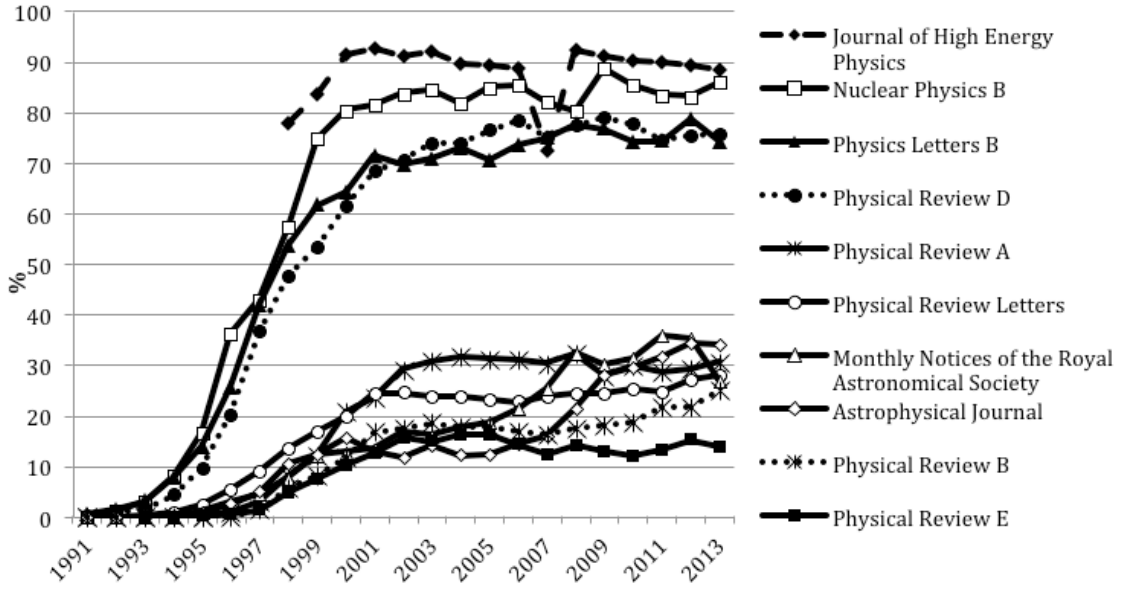

Figure 4: Time series of the percentages of papers in a journal that cite arXiv. Displayed are the 10 journals that most actively cite arXiv.

Striking is the decline of the share of papers in JHEP with references to arXiv in 2007, for which no explanation can be given. Overall, the shape of the line graphs suggests a rapid growth of arXiv's acceptance in the 1990ies and a constant reliance on arXiv in the past 15 years. The following table identifies other "arXiv-friendly" journals, where the majority of papers rely on arXiv. Since the number of papers published in a journal can differ immensely, Table 5 indicates percentages of the number of a journal's papers that cite arXiv. To provide an up-to-date view, only papers published between 2004 and 2013 are considered.

Table 5: Journals in WoS with the highest share of papers citing arXiv. Analyzed are only citing papers that were published between 2004 and 2013.

Journal	%	Journal	%
Journal of Cosmology and Astroparticle Physics	89.9	Journal of Physics G-Nuclear and Particle Physics	59.0
Advances in Theoretical and Mathematical Physics	81.7	International Journal of Modern Physics A	59.0
Annual Review of Nuclear and Particle Science	80.7	International Journal of Modern Physics D	57.5
Communications in Number Theory and Physics	79.8	Progress of Theoretical and Experimental Physics	56.2
European Physical Journal C	70.9	Physics Reports-Review Section of Physics Letters	55.5
Fortschritte der Physik-Progress of Physics	70.4	General Relativity and Gravitation	54.0
Quantum Information & Computation	69.3	Gravitation & Cosmology	54.0

Modern Physics Letters A	62.3	Journal of Sympletic Geometry	53.6
Progress in Particle and Nuclear Physics	61.5	Reviews of Modern Physics	52.8
Acta Physica Hungarica A-Heavy Ion Physics	60.4	Algebraic and Geometric Topology	52.2
Geometry & Topology	60.3	Progress of Theoretical Physics	51.7
Classical and Quantum Gravity	60.0	Astroparticle Physics	51.2

Ranking the journals on the basis of percentages instead of absolute numbers enables us to spot mathematics journals. The 24 journals listed prove that the circle of users coincides with the target group of arXiv that consists mainly of high-energy physicists. In HEP it is usual practice to submit papers to arXiv prior to journal submission. According to Gentil-Beccot, Mele & Brooks (2009) the arXiv often presents a version very similar to the published one. Finally, the arXiv version is freely available, while the journal versions require subscription.

Conclusions

The rapid dissemination of research results enabled by arXiv has accelerated the read-and-cite process (see Brody, Harnad & Carr, 2006). The identified number of cited references to arXiv and the rapid citation of e-prints in WoS-indexed journals indicate that e-prints are accepted within certain communities as well as among journal editors. Taking citation counts as a proxy for quality, e-prints on arXiv can be regarded as of good quality. They are valued, read and used within the scientific community, mainly because they present results upon finalization, circumventing the publication delay. To refer to these most up-to-date findings, authors evidently do not hesitate to cite arXiv e-prints in their research papers. The high number of cited references presented in this study suggests the usage of e-prints over the journal articles, as it was also found by Davis & Fromerth (2007). One reason for the preference of arXiv e-prints is the free availability of full text, especially if readers do not have access to the journal. Besides, the arXiv version is often similar to the formal journal article and can be easily cited by ID. An obvious reason to cite arXiv full texts even years after publication might be simply that the e-print does not have a published alter ego to be cited. Furthermore, the results showed that citations to e-prints peak in the year after publication and drop rapidly in the following years. Authors may still rely on the e-print but cite the formal publication, so the decline in citations does not necessarily indicate a decline in use. This could be proved in a future study with download data of arXiv e-prints over time.

Whereas this initial study is mostly exploratory, future work will link arXiv data to the data in WoS to examine, whether the cited e-prints have a journal version or not. So far, Larivière et al. (2014, p.1161) found that 64% of all arXiv e-prints are published in a WoS-indexed journal. An improved unification in our bibliometric database of institution names will allow analysing reasons why certain institutions rely on arXiv. Is it due to the presence of large physics departments, research centres, outstanding and highly-active researchers, collaboration or cutting-edge research? Moreover, a qualitative study of authors and their reasons to cite arXiv instead of the journal article would provide valuable information on the recent scholarly communication process.

References

Amat, C. B. (2008) Editorial and publication delay of papers submitted to 14 selected Food Research journals. Influence of online posting. *Scientometrics* 74, 3, 379-389.

Antelman, K. (2004). Do open-access articles have a greater research impact? *College and Research Libraries*, 65 (2004) 372 – 382.

Brody, T., Harnad, S., & Carr, L. (2006). Earlier web usage statistics as predictors of later citation impact. *Journal of the American Society for Information Science & Technology*, 57(8), 1060–1072.

Brown, C.M. (2001). The E-volution of preprints in the scholarly communication of physicists and astronomers. *Journal of the American Society for Information Science and Technology*, 52(3), 187–200.

Brown, C. (2003). The role of electronic preprints in chemical communication: Analysis of citation, usage, and acceptance in the journal literature. *Journal of the American Society for Information Science & Technology*, 54(5), 362–371.

Davis, P.M., & Fromerth, M.J. (2007). Does the arXiv lead to higher citations and reduced publisher downloads for mathematics articles? *Scientometrics*, 71(2), 203–215.

Garner, J., Horwood, L., & Sullivan, S. (2001). The place of eprints in scholarly information delivery. *Online Information Review*, 25(4), 250–253.

Gentil-Beccot et al. (2009). Information Resources in High-Energy Physics: Surveying the Present Landscape and Charting the Future Course. *Journal of the American Society for Information Science and Technology*, 60 (2009) 150–160.

Gentil-Beccot, A., Mele, S., & Brooks, T.C. (2009). Citing and reading behaviours in high-energy physics. How a community stopped worrying about journals and learned to love repositories. Retrieved January 6, 2015 from: arXiv:0906.5418.

Ginsparg, P. (1994). First steps towards electronic research communication. *Los Alamos Science,* 22, 156-165.

Goldschmidt-Clermont, L. (1965). Communication Patterns in High-Energy Physics. Retrieved January 6, 2015 from: http://eprints.rclis.org/archive/00000445/02/communication_patterns.pdf

Haque, A., & Ginsparg, P. (2009). Positional effects on citation and readership in arXiv. *Journal of the American Society for Information Science and Technology*, 60(11), 2203–2218

Haque, A., & Ginsparg, P. (2010). Last but not least: Additional positional effects on citation and readership in arXiv. *Journal of the American Society for Information Science & Technology*, 61(12), 2381-2388.

Harnad, S. & Brody, T (2004). Comparing the Impact of Open Access (OA) vs. Non-OA Articles in the Same Journals*, D-Lib Magazine* 10.

Henneken, E.A. et al. (2007). E-prints and journal articles in astronomy: A productive co-existence. *Learned Publishing*, 20, 16-22.

Heuer, R.-D., Holtkamp, A., Mele, S. (2008). Innovation in Scholarly Communication: Vision and Projects from High-Energy Physics. pp.1-15. DESY-08-054. Retrieved January 6, 2015 from: https://bib-pubdb1.desy.de/record/86123/files/getfulltext.pdf

Kurtz, M & Henneken, E. (2007). Open Access does not increase citations for research articles from The Astrophysical Journal, Retrieved January 6, 2015 from: arXiv:0709.0896

Langer, James. (2000). "Physicists in the new era of electronic publishing." *Physics Today Online*, 53(8):35-38.

Larivière, V., Sugimoto, C.R, Macaluso, B., Milojevic', S., Cronin, B. & Thelwall, M. (2014). arXiv E-Prints and the Journal of Record: An analysis of Roles and Relationships. *Journal of the American Society for Information Science & Technology*, 65(6):1157–1169.

Lim, D. (1996). Preprint Servers: A New Model for Scholarly Publishing? *Australian Academic and Research Libraries (AARL)* 27 (1), 21–30.

Luwel, M. (1998). Publication delays in the science field and their relationship to the ageing of scientific literature. *Scientometrics*, 41, 29-40.

Metcalfe, T.S. (2005). The rise and citation impact of astro-ph in major journals. *Bulletin of the American Astronomical Society,* 37, 555–557. Retrieved January 6, 2015 from: http://arXiv.org/abs/astro-ph/0503519

Metcalfe, T.S. (2006). The citation impact of digital preprint archives for solar physics papers. *Solar Physics,* 239, 549-553.

Moed, H.F. (2007). The effect of 'Open Access' on citation impact: An analysis of arXiv's condensed matter section. *Journal of the American Society for Information Science & Technology*, 58(13), 2047–2054.

Prakasan, E.R. & Kalyane, V.L. (2004). Citation analysis of lanl high energy physics e-prints through Science Citation Index (1991-2002). Retrieved January 6, 2015 from: http://eprints.rclis.org/archive/00002200/

Schwarz, G.J., Kennicutt, R. C. J. (2004). Demographic and citation trends in astrophysical journal papers and preprints. *Bulletin of the American Astronomical Society,* 36 (2004), 1654–1663.

Shuai, X., Pepe, A. & Bollen, J. (2012). How the Scientific Community Reacts to Newly Submitted Preprints: Article Downloads, Twitter Mentions, and Citations. *PLoS ONE* 7(11): e47523.

Youngen, G.K. (1998). Citation patterns to traditional and electronic preprints in the published literature. *College & Research Libraries*, 59(5), 448–456.