

THE EXPERIENCE OF THE REPEC PLAGIARISM COMMITTEE IN ECONOMICS

Christian Zimmermann

Abstract:

RePEc is an open bibliography project driven entirely by volunteers and without a budget. It was created to enhance the dissemination of research in economics by making it more accessible to authors, publishers, and readers: 1800 publishers participate in this initiative, and 44000 authors are registered.

Some of those authors became frustrated when their work was plagiarized and no action was taken. Many have asked whether RePEc could take action. The RePEc Plagiarism Committee was created to respond to this request. Because RePEc has no enforcement power, it can only “name and shame” verified offenders. This essay discusses the experience over the first years of the Committee.

Key words: Plagiarism; academic literature; naming and shaming; economics

1 RePEc

RePEc stands for “Research Papers in Economics.” It was created in 1997, following some precursor initiatives dating back to 1992, to enhance and democratize the dissemination of research in the field of economics. The field suffers from extensive publication delays; an article can take many years to pass through the peer-review process. Hence, an active pre-print culture has evolved. Before the widespread use of the web, though, it was very difficult to obtain these pre-prints if one did not have access to the networks that allowed such sharing. This meant that the frontier of research was accessible only to economists in elite institutions. Others discovered the research only years later after publication in journals.

RePEc changed this system by allowing any institution to contribute its publications for indexing in RePEc at no cost. The dissemination, which is also free, has been performed by volunteers through websites, mailing lists, and social media. Eventually, RePEc grew both in content and popularity to the point that even commercial publishers started volunteering metadata for their publications. Finally, other initiatives for the economics profession have grown out of RePEc thanks to its central location in the community of researchers.

Engagement in RePEc expanded significantly once it started publishing statistics about authors and their institutions, first based on user downloads of full texts through RePEc services¹ and then by incorporating citation analysis of indexed works. This publication of statistics has allowed the establishment of rankings and has motivated authors to ensure all their works are listed. Authors have exerted pressure on their home institutions and publishers to have their publications listed in RePEc. At the time

¹The most popular RePEc services are IDEAS (<https://ideas.repec.org/>) and EconPapers (<http://econpapers.repec.org/>), which both allow searching and browsing of material indexed in RePEc, and NEP (<http://nep.repec.org/>), which disseminates new pre-prints through email, RSS feeds, and Twitter.

of this writing, about 1800 publishers are participating, covering over 2000 journals and 4000 pre-print series, as well as software, books, and book chapters, for a total of 1.7 million documents.

2 The Motivation for the RePEc Plagiarism Committee

There are four premises that led to the constitution of the RePEc Plagiarism Committee. The first is probably not unique to economics: It is the lack of response to formal complaints of plagiarism filed with the publisher or institution the presumed plagiarizer. It is very frustrating when any author discovers they have been plagiarized and files a complaint where the offender has published, but nothing happens. Without other options, several authors have turned to RePEc asking for help. At first, with no process in place, there was little RePEc could do.

The second premise is that even when a plagiarizer is punished through a retraction and/or sanction at the home institution, the knowledge of this event is usually contained. There have been cases where an offender has been fired and then found new employment where no one was aware of the plagiarism. It is not unheard of that an offender relapses and plagiarizes again.²

The third premise is specific to economics and RePEc in particular: RePEc has made it easier for plagiarism to be discovered. Because it is in the interest of authors and publishers to list all their work, even articles in obscure journals become discoverable. The most spectacular case was a journal that was publishing articles without the authors' consent and under their true names. The authors soon discovered this; the journal was banned from RePEc, and it returned to obscurity. Surprisingly, many plagiarizers contribute the work they have plagiarized in RePEc, to be listed publicly.

This brings us to the fourth premise, which is also specific to RePEc. The publication of rankings has generated much RePEc's growth, but it also has motivated some authors to boost their scores in unethical ways. One way has been to cheat by doctoring their download numbers, something that was quickly dealt with through improved counting algorithms and occasional sanctions. This task of computing the statistics is performed by RePEc volunteers independently of the RePEc Plagiarism Committee. The other instance of cheating is unethically increasing the numbers of works listed in an author's RePEc profile, for example through plagiarism.

Before the Committee was constituted, several spectacular cases of plagiarism came to light. At the time, all what RePEc could do was to remove those works. Even though home institutions were notified, they often did not react, likely because the complaint was not coming from a source they considered impartial or capable of imposing consequences.

Note that in some cases the RePEc Plagiarism Committee responded and took steps that publishers could have taken in trying to enforce their own copyrights. They may not have had a sufficient commercial interest to do so, but the ethical interests of science may be stronger. Thus, the Committee sees it as its mission to pursue cases of plagiarism even if publishers do not.

²Unfortunately, it is not possible to be more specific about these cases because they have not been made public and have not been formally scrutinized by the RePEc Plagiarism Committee.

Here is an example. A graduate student posts a large number of research papers on RePEc within a relatively short time. Of course, this raises suspicions, and reports soon start coming in from authors that they have been plagiarized *in extenso* from other works already listed in RePEc. The student has only changed the title, the author, and the references, introducing multiple references of his own (plagiarized) works. All this was quickly detected and his graduate program was contacted. The administrators of the offender's graduate program promise to address this and soon they communicate by email that the offender was expelled from the program. A few years later, RePEc volunteers look into the case and find that the offender graduated with a doctorate from the very same program. This was not an obscure graduate program, and it was in a country with a good reputation for research ethics.

While the example above may be the most spectacular, it also highlights how difficult it can be to enforce ethical behavior and impose penalties in any such cases. So what is one to do? The approach that was chosen was to “name and shame” offenders on a public website. This means that once a clear case of plagiarism has been determined, all documents relating to it are made public and the offender is named. The naming is crucial here. First, it should discourage future potential offenders. Second, it still imposes consequences on those offenders who are not penalized by their home institutions. Third, in the case that offenders seek new employment, they cannot hide their history.

3 Constitution and Rules of the Committee

Accusations of plagiarism are a serious matter, even more so if offenders are to be named and shamed. The procedures thus need to be rock solid and decisions should be clear. The Committee making those decisions should be beyond reproach and should adhere strictly to procedures. The Committee should also be diverse in representing different regions, institution types, and fields of research.

A call for members was disseminated through the RePEc Blog and through the monthly mailing to authors registered with RePEc. The goal was to enlist 20 to 25 members. This large number is intended to guarantee wide coverage, including multiple areas of research and expertise, and protect members from pressures that could occur if the Committee were much smaller. The broad spectrum also allows the Committee to be free from external pressures—for example, from national professional associations, publishers, and other institutional bodies.

The founding Committee was a mixture of economists who are already active volunteers in the RePEc community, others who have some experience with plagiarism (either as victim, editor, or researcher), and newcomers. The first task was to establish the rules and procedures.

The Committee decided it is willing to address cases of plagiarism that have not yet been publicly documented, say, through a public retraction. For cases that have already been made public, the Committee determined it is sufficient to list these offenders on the Committee's website. Second, for a case to be studied, one of the involved works needs to have been listed on RePEc. It is not the goal of the Committee to cover other fields of study or other public forums, and the restriction to RePEc provides the right

scope and boundaries. It is, however, clear that not all research in economics is covered by RePEc, but one can make the case that most of it is. Third, a case can be submitted to any Committee member, who is required to do some initial vetting to see whether the case has enough credibility and substance. Later, it was also decided that anonymously submitted cases would not be accepted. The main concern here was that anonymous submissions could be motivated by other factors (including vendettas or harassment) well beyond the Committee's mission. In any case, the identity of submitters is not made public, even to the other Committee members.

Once a case is accepted for consideration, the accused offenders have the opportunity to defend themselves. After the case is assembled, it is sent to the accused offender and they have two weeks to respond to the charges. Thereafter, everything is submitted to the Committee, who discuss the case and vote anonymously.

It was decided that for any decision to be reached, a two-thirds majority of current Committee members is necessary. This is a very high hurdle. First, it is more than the simple majority required in most decisions. Second, any abstention counts as a no. Being so strict was felt to be necessary given the potentially strong implications of a decision.

How does the Committee vote? It is not sufficient to determine whether there is plagiarism. There is also the decision about what to do about the plagiarism. The Committee votes separately on the following decisions:

- Alert affected author;
- Alert affected editor;
- Alert administration at home institution of offender;
- Exclude author from RePEc;
- Request removal of plagiarizing work;
- Make the case public.

One may wonder whether excluding the offending author from RePEc has any impact. It can in some cases, as the publication then cannot be applied to the home institution in the rankings of institutions. It was also later decided to give the option to keep the plagiarized work on RePEc, but request the full text to be watermarked to indicate the offense.

Finally, the Committee does not strictly state that plagiarism is present. By alerting some people or making the case public, it makes it possible for others to form their own opinion. But the Committee makes public only clear and credible cases.

4 The Committee's Work So Far

After establishing the rules and procedures, the Committee started taking its first cases in February 2011. Initially, the work was perceived as rather light. Reasons could be that relatively few people were aware of the Committee and did not submit cases, that the reputation of the Committee had not yet been built and authors were hesitant, and that the expected backlog of plagiarism cases just did not materialize. Over time, the number of cases increased. The first cases were rather clear-cut, but some of the later

cases were not. For example, a few cases had some self-plagiarism mixed with typical plagiarism or evidence that was not very clear. Obviously, judging a plagiarism case is more difficult when it is not straightforward. Yet, a very large percentage of cases has been decided with very clear majorities.

The most difficult cases turned out to be those where ideas, and not words, were supposedly plagiarized. In such cases, it is not sufficient to compare text; one needs to actually unravel and understand the concepts. In one case, it was necessary to call for outside referees more familiar with the research area, which turned out to be very difficult. For one, referees are used to contributing peer-review for publication, but not for conflicts. Second, in highly specialized areas of research, few economists do not have some conflict of interest with one of the involved parties.

Surprisingly, some cases never made it to the Committee because the aggrieved parties were not able to provide enough evidence. One example involves an economist alleging that another economist used a presentation of his delivered at a conference to quickly write a paper and submit it before his draft was completed. The accuser was, however, never able to supply his initial presentation.

Another victim and accuser decided to withdraw the case. The case was not crystal clear, as it implied a plagiarism of ideas. The accuser was also afraid of adverse consequences as the accused economist was rather prominent in the field. This result was unfortunate, but the Committee should not overrule the decision of the accuser.

Normally, votes are almost always very clear one way or the other. But the Committee has found it difficult to reach a decision in some cases because it set itself a very high hurdle with the $2/3$ majority of current members. The difficulty is mainly in achieving a sufficient number of votes.

At the time of this writing, the Committee has publicized 20 cases on its website at <https://plagiarism.repec.org/>. The reader can independently judge the merits of those cases. Note that there are repeat offenders, although so far no one has repeated after being incriminated by the Committee. In some cases, offenders have had high standing in their home institution thanks to their research, some of which turned out to be “borrowed.”

The Committee website also lists 25 plagiarism cases that have been independently documented, typically through explicit retractions. This additionally includes 12 cases of self-plagiarism (publishing the same work in several journals) and 22 cases of other unethical behavior (manipulation of data, for example). Overall, 78 authors are named and shamed, 12 for multiple cases. This listing further highlights the fact that some repeat offenders in the profession should have been named long ago.

Judging from informal conversations and observations of social media, economists are now very much aware of the existence of the Committee. They do look at the published cases, to the point that papers involved in plagiarism sometimes receive the most downloads in RePEc. This has the perverse effect that offenders may actually receive a boost in their download statistics from the RePEc services that are used for rankings.

One should also add that the existence of the Committee has raised awareness about plagiarism in the profession. Published cases have led to the adoption of plagiarism policies in affected institutions and elsewhere in an attempt to avoid being listed on the

website. One institution even volunteered a case before the Committee had learned about it, asking that the name of the institution not be mentioned before it could take disciplinary action.

5 Looking ahead

We hope that the existence of the RePEc Plagiarism Committee will curtail the incidence of plagiarism. Such unethical behavior is, however, unlikely to disappear, especially as current technology makes it easier to copy other work. A particular challenge that may lie ahead is dealing with cases of unauthorized translation without attribution. Such cases are very difficult to detect using automated screening processes. Also, victims are unlikely to discover such offending works.

Another class of cases that could be challenging are those dealing with self-plagiarism. There is no clear agreement, neither in the profession nor in the Committee itself, on how to define its boundaries. So far, the website lists only those cases documented elsewhere.

Copyright statement

Copyright © 2015. Author(s) listed on the first page of article: The author(s) grants to the organizers of the conference “Plagiarism across Europe and beyond 2015” and educational non-profit institutions a non-exclusive licence to use this document for personal use and in courses of instruction provided that the article is used in full and this copyright statement is reproduced. The authors also grant a non-exclusive licence to Mendel University in Brno, Czech Republic, to publish this document in full on the World Wide Web (prime sites and mirrors) on flash memory drive and in printed form within the conference proceedings. Any other usage is prohibited without the express permission of the author(s).

The author thanks members of the RePEc Plagiarism Committee for comments on a previous draft. The views expressed are those of individual authors and do not necessarily reflect official positions of the Federal Reserve Bank of St. Louis, the Federal Reserve System, or the Board of Governors.

Author

Christian Zimmermann (zimmermann@stlouisfed.org), Federal Reserve Bank of St. Louis, Economic Research, P. O. Box 442, St. Louis MO 63166, USA