

## Text recycling

Self-plagiarism, also referred to as 'text recycling', is a topical issue and is currently generating much discussion among editors. Opinions are divided as to how much text overlap with an author's own previous publications is acceptable, and editors often find it hard to judge when action is required. In an attempt to get some consensus and consistency on the issue, editors at **BioMed Central** have produced some guidelines. These guidelines were posted on the COPE website and members were invited to comment.

The topic was discussed at the COPE Forum on 12 March 2013.

### **COMMENTS FROM THE FORUM (12 MARCH 2013) – NOTE, Comments do not imply formal COPE advice, or consensus.**

- Is the term text recycling better than self-plagiarism? Should we stop using the term self-plagiarism?
- Text recycling is a better term (than self-plagiarism) because it is very important to separate true plagiarism from repetitive reporting of your own work. Most institutions regard true plagiarism and text recycling as two very different types of misconduct. True plagiarism is seen by most people as a serious form of misconduct whereas text recycling is sometimes just inevitable. Text recycling may also be quite legitimate (eg, in the methods section) and so is not misconduct. Text recycling is a neutral term. Any guidelines need to be specific and include guidance just on publications but on other forms of publication, particularly on the web.
- We should continue to use the term self-plagiarism as many authors do not think this is an unethical practice. Text recycling implies it is acceptable.
- This is an issue for editors—to separate the words from the content. So the words might be the same but the data should not be the same. The data should be new, should contribute to the literature and should not be just a repackaging of the same published material. This is the job of editor to sort out.
- Would it be useful to have percentages of the amount of text recycling that is allowable as a guideline, similar to copyright law (which is 10%)? This would not apply to data, but maybe to other sections (eg, methods)?
- Many editors and editorial staff put too much emphasis on percentages, and what percentages they should be looking at. We should abandon this practice as it can be very misleading—1% can be really damaging for instance. Editors should consider each case of text recycling on an individual basis (as stated in the guidelines). Editors should eyeball all papers.
- There is a lack of knowledge, especially among young researchers, of what can be classified as prior publication, particularly in relation to online publishing, conference proceedings, etc. Journal policies often vary on this issue. So this could be incorporated into this type of documentation.
- The issue of text recycling is different in different disciplines— text recycling is more acceptable in some areas than in others so we need to recognise the differences between academic fields.
- Salami slicing to some extent overlaps with text recycling.

- Where text recycling occurs in the document is also important. While we might encourage text repetition in materials and methods, have some sympathy for text recycling in the introduction but very little in the discussion, text recycling in the results is completely unacceptable. So issue of where it is appropriate and where it isn't should be discussed.
- There is a spectrum of behaviours, from acceptable to misconduct, a large part of which can be attributed to author laziness (recycling in the introduction, for example).
- We should proactively encourage authors to be transparent, creating a transparent environment.
- What about papers where there are multiple authors and you share authorship? How do you attribute that? How does one reuse the text of methods when authors differ from one paper to the next?
- Text recycling as a term implies that it is acceptable, are we comfortable with that view? Is text recycling a good thing to do? (On a show of hands, about third of the Forum audience said they thought text recycling is a good thing and should even be encouraged.)
- There may also be legal issues related to copyright infringement.
- Referencing previous work is absolutely essential and should be stressed, making it clear where the original text came from.
- There was a discussion about the 2004 cut-off date chosen by BioMed Central. BioMed Central said it was an arbitrary choice taking into consideration that self-plagiarism has been made more easy to detect in the past decade.
- If you do have a cut-off time, it will make it easier in terms of dealing with anonymous whistleblowers and cases of redundant publications, but this needs to be accompanied by clear guidance (possibly even included in the instructions to authors), saying we know this is arbitrary, we know it was a problem before this date but it is simply a practical cut-off point.
- A big problem in both text recycling and conventional plagiarism is the remedy. When authors are asked to correct duplicate text, the typical result is proforma substitution of words. The basic text then remains with "new" words interspersed here and there. This addresses the letter of the law rather than the spirit of the law as it were. Frequently the resulting sentences are more awkward than the original. Getting a satisfactory response is then very time consuming.

**Action:** BioMed Central will revise their guidelines in response to all of the comments, both at the Forum and on the website, and incorporate them into their policies and instructions for authors. They would also be interested in feedback on whether this should be applied differently to review articles? After discussion with BioMed Central, the guidelines may also be posted on the COPE website stating that they were developed by BioMed Central in association with the COPE Forum.

#### **COMMENTS POSTED ON THE WEBSITE**

*Posted by Charlotte Seidman, 22/2/2013 7.42pm*

Thanks for focusing the spotlight on this important topic. I've been "iThenticating" all revised AJPM papers for several years now, and am continually frustrated by self-plagiarism. I now have two lines that I repeat to our editors on a regular basis: "Self-plagiarism is, by its very name, plagiarism"; and "you'd think that researcher/authors with MDs and PhDs would be bright enough to know how to reword." Levity aside, I'm pleased that COPE has created guidelines for us. Thanks!

*Posted by Justin Starren, 28/2/2013 2.03pm*

In large collaborative projects that generate multiple articles, there is a need for concise introductory paragraphs that describe the larger project in order to put the narrower article in context. To leave room for the real content, such paragraphs need to be tightly worded, and may be refined by multiple authors. Forcing authors to reword such paragraphs does nothing to improve the content of the article. Typically it only serves to lengthen the article and reduce the quality of the writing.

*Posted by Charlotte Seidman, 22/2/2013 7.43pm*

Thanks for focusing the spotlight on this important topic. I've been "iThenticating" all revised AJPM papers for several years now, and am continually frustrated by self-plagiarism. I now have two lines that I repeat to our editors on a regular basis: "Self-plagiarism is, by its very name, plagiarism"; and "you'd think that researcher/authors with MDs and PhDs would be bright enough to know how to reword." Levity aside, I'm pleased that COPE has created guidelines for us. Thanks!

*Posted by Reid Lifset, 22/2/2013 7.51pm*

It would be very helpful if there were some discussion of situations where text recycling occurs when some but not all authors are the same for the papers in question. Also, the guidance should indicate that the authors are obligated to disclosure text recycling to the editors, that is, that the burden of disclosure rests with the authors. Finally, some discussion as to when and how "clarifying what is new in the subsequent publication versus the original publication" is appropriate would be helpful as well.

*Posted by Steven G. Silve..., 22/2/2013 7.58pm*

In my field (Pathology), and probably in others as well, a much bigger problem than recycled text (probably not so bad in an invited review article, in which the authors are specifically invited to discuss their previous publication/s, obviously with appropriate citation) is recycled photographs, diagrams/drawings, Tables and the like. Again, the rule we usually follow is to require permissions from the original publisher/s and, obviously, citations. I must say, however, as an author as well as an editor, that it has always annoyed me to have to get permission from a publisher to reuse my own original illustrations. Any comments?

*Posted by Peter LaPlaca, 22/2/2013 9.14pm*

Another type of self-plagiarism is when authors try to milk a research project to get multiple publications, each with a slightly (in some cases, very slightly) different angle. Assuming proper citations to the previously published articles, my general remedy is to apply a contribution metric to the paper: if the marginal contribution of the new paper is significant, I will ask the authors to indicate in their introduction that this is a "continuation" paper of a previous article and send it out for review; if the contribution is negligible, I will reject the submission.

*Posted by Chris Barrow, 22/2/2013 9.15pm*

This is a significant problem and growing I think.....amazingly some argue with Crosscheck scores. A related issue is the author who multi-submits in spite of cautions not to - and you find

more than one journal refereeing essentially the same (possibly then accepted) papers....if the journals are in a range of countries it may be difficult to spot

*Posted by Mohammad Abdollahi, 23/2/2013 5.33am*

As editor I have the experience of working with Cross Check. Cross Check reports similarity index and use of similar words. The point is that report of similarity is not enough to judge about the intention of authors to copy-paste or plagiarism. I mean there are sometimes limited words specifically scientific words to state something in English and if one searches Cross Check will find almost all forms of words in previous papers. So use or interpretation of Cross Check report is something that should be learned by editors. Just having a similarity index of 30% for instance cannot mean anything if editors do not go through whole paper. In fact such similarity is not restricted to Methods section. For instance, there are limited forms to state Results in scientific papers. I fear to reach somewhere to lack enough English words to state something to be absolutely different from previous papers. This is what we have to be careful and think about.

*Posted by Graham Pawelec, 23/2/2013 6.07am*

I think that one has to distinguish between original articles and review-type articles when applying the detailed section-by-section which is absolutely necessary for each analysis of similarity reports. In the former, the most important issue is to prevent data duplication. Materials and Methods commonly show duplication, acceptable under most circumstances even if including data which might also sometimes be reported in the Results section (eg, clinical characteristics of patients used in several studies). Significant new data in the Results will always require new text for Discussion. If not, the paper was probably not novel enough to warrant publication anyway. Thus, my tolerance for self-plagiarism is highest for M&M, next highest for the Intro, low for Discussion and zero for Results. In review-type articles, especially those by non-native speaker writers, some guidance on reformulating text passages may be required, but "recycling" chunks of text is not such a sin here, in my opinion, as long as there is no hidden duplication of the entire paper.

*Posted by Roger Brown, 23/2/2013 9.23am*

From the viewpoint of Polymer Testing, the important area is the results. There must be significant or important new results but it is often reasonable to reproduce descriptions of test methods, materials and even parts of the introduction. If the test method used is not new it must be made very clear where it originates and why it is used..

*Posted by Philip Scarf, 23/2/2013 9.32am*

A growing issue in my field is the practice of publishing a paper in a conference proceedings and then extending this paper for publication in a journal. There is often considerable overlap between these papers. I do not think authors are being disingenuous. They are typically being encouraged by conference organisers and editors to follow this practice. Conference proceedings are often considered to be valuable by the researchers in the field themselves (as they collect together a large cognate and current body of knowledge) but papers in conference proceedings are not considered to be valuable in research assessment exercises (as the crude view of research assessors is that conference proceedings are not subject to the same standards of peer review as journals). So, an author risks accusations of plagiarism (albeit text recycling) in

order to satisfy these two masters: science and career. I personally would be very unhappy to see some of the suggestions above (e.g. retraction, publication of a correction) implemented in this kind of case. However, that said, I am uncomfortable with the proceedings-journal "double" publication practice.

*Posted by Ian W Henderson, 23/2/2013 1.38pm*

A sometimes trite sentence or remark is surely acceptable, but wholesale repetition of one's own prose is frankly boring. Repetition of data, without due acknowledgement, cannot be used. Some opinions on their publication are sometimes draconian especially if there is only a suspicion of self-aggrandisement. More seriously, does the self-plagiarism amount to pure duplication? In a learned scientific publication self-plagiarism of solid data, without acknowledgement is unacceptable. In other circumstances, there is often no (or but a few) other ways of giving the clear information in prose form - this may apply to all literature, of course.

*Posted by John Kleinig, 23/2/2013 6.22pm*

It is much easier to acknowledge the problem than to know how editorial discretion is to be exercised. It would probably be wise to revise the text of the document in light of the various comments submitted, noting in more detail the kinds of issues/factors that might lead to different actions being taken (there is surely no "one size fits all" solution). Greater author transparency as to the degree of overlap when a previous article is used would also be helpful. I do, however, wish to protest against the language of "self-plagiarism" -- the point is not that one is represent others' ideas as one's own, but that one is recycling one's own work. That is a different fault.

*Posted by Jacob L. Mey, 23/2/2013 8.57pm*

A common scenario is this: one presents a paper at a foreign university, then is asked to publish it in their local journal or collection of working papers. As an instance, I have several papers in local publications of Japanese universities (some in Japanese); the same is true for Brazil. Not many people I know have access to these sources and I regularly republish such papers (translated if needed) in more internationally accessible journals. I feel OK with this practice, esp. as I always refer to the original place of publication, thus giving my foreign colleagues a small leg-up.

*Posted by Fatema Jawad, 24/2/2013 4.24pm*

It has been observed in our journal, that authors conduct one study and the results are divided in two portions and submitted with a small difference in the results achieved. In such a case there is overlapping in nearly all sections. It actually falls in the section of Salami slicing with self plagiarism or Text Recycling. Our editorial board members consider this unethical and following the guidelines of COPE we ask for an explanation. Usually the answer is not convincing and we do not process the article. It is considered rejected on the grounds of ethical misconduct.

*Posted by Christopher Moran, 24/2/2013 11.31pm*

I agree with the comments above about the location of the self plagiarism within the paper being critically important eg from M&M - sometimes unavoidable, to results - unacceptable. My other concern though is how the plagiarism software actually work. I was dumbfounded to

discover that the package provided by my journal, which I only rarely use, identified the references in the bibliography as "plagiarised". When I discounted this, a manuscript with a very suspicious score initially had quite benign levels of apparent "plagiarism", requiring no further action.

*Posted by Michael Wise, 25/2/2013 1.21am*

This is a more subtle problem than plagiarism from other sources. The proposed Guidelines do strike the right note. My only suggestion is that it would be helpful to operationalise the Guidelines by saying that, say 10% duplication in Introduction and Methods sections (but not Results) should not normally be cause for comment. BWT, (to Chris Barrow), based on my use of other, similar systems, I wouldn't rely solely on Crosscheck scores, but also eyeball the extent of the actual match. Specious matches can occur, e.g. including a sequence found in a database.

*Posted by Richard Saitz, 25/2/2013 4.07am*

The guideline might state when text *should* be recycled. Text should be recycled when a paper is reporting different findings from a study previously reported. Why should synonyms be used for methods? If the methods were identical they should be described using the same words. This might apply to papers reporting subgroup analyses from clinical trials or observational analyses of trial data. Similarly, when describing participant characteristics in those cases, hopefully they are the same, if the sample was the same. Using synonyms and mixing up words serves no purpose in those cases. On the other hand, there is clearly duplication that should not happen, like reporting the same discussion, or analytic results. Introductions could present similar arguments but would be unlikely to be justifiable as identical.

*Posted by Georgii Alexandrov, 25/2/2013 8.13am*

The use of such term as 'self-plagiarism' makes false impression that 'text recycling' is a kind of plagiarism. This is not the case. It is legal under some circumstances to insert a section of text from a previous article into a new one. The problem is that authors do not inform editors about the 'recycled' sections. I suggest to avoid the term 'self-plagiarism' and to add the phrase, "Editors may ask authors to confirm that there are no recycled sections in the submitted manuscript or to provide the list of such sections".

*Posted by Lars Ole Sauerberg, 25/2/2013 10.53am*

In the humanities and the social sciences, where book publications have had, and still enjoy, prominence, the dynamics of recycling are particularly to be seen 1) in the interaction between book and paper/article publication, with books chapters redesigned as papers/articles and vice versa, and 2) in the effort to broadcast results originally relating to language/culture contexts different from English to a larger audience by way of translation into English. As long as the publication history is indicated clearly, both to editors and readers, there should be no problem in this.

*Posted by Phil Wiffen, 25/2/2013 1.35pm*

Self plagiarism is an oxymoron. I work in systematic reviews and I would prefer to see the same correct methods used across a number of SRs than in correct ones. There are not many ways to describe how bias was identified for example and I am involved in 10-20 new reviews a year. Are

some of the commenters to this list wanting every one of these to be totally different??  
Obviously this doesn't apply to results etc

*Posted by Lin Perry, 26/2/2013 10.19am*

Thanks for thoughtful guidelines. This is a discussion that will run - terms such as 'small', minor' and 'major' provide space for the differing interpretations reflected in the comments above in relation to the range of subject fields. They also offer a wide open barn door for subjective and personal interpretation - which will mean author challenges. Pure metrics aren't the answer though, as WHAT has been duplicated is as important, if not more than HOW MUCH. I don't see a quick answer: more that increasing transparency and greater discussion of the issue will lead to (perhaps) increasing consensus and better informed/ prepared authors. Well, we can hope!

*Posted by Julie Browne, 27/2/2013 1.02pm*

I don't see what's wrong with authors writing "The methods used in this study were previously described xxx (article reference)" thereby indicating who was the originator of the text and freeing up publication space for genuinely original work. Such an approach would prevent the not uncommon situation where the authors of several papers containing recycled text are not all the same people, or their names appear in a different order, or the corresponding author is different from paper to paper. On one occasion in my experience, quite large sections of text were being passed around between a number of people and papers and it became unclear (certainly to us editors, but also, I think to the authors) whose text it actually was in the first place. It was beginning to blur into genuine plagiarism, except that we could not untangle the paper trail.

*Posted by David Baxter, 27/2/2013 7.27pm*

I wondered whether are any implications in respect of publications arising from doctoral theses, which are increasingly being archived in open access repositories by universities. This may well be a wider issue, but text recycling would seem to be unavoidable in such cases, and perhaps worthy of some mention or specific caveat in published guidelines?

*Posted by Robert B. Stevenson, 1/3/2013 11.44pm*

Some medical journals like JAMA and NEJM will not accept articles with previously published information. Dr. Arnold Relman, NEJM editor explained his readers wanted new, fresh information, not a rehash. I agree, and recommend searching submissions for previous publication of text as well as photos, if only to avoid copyright problems. A related topic is when a single research study is conducted by four individual co-authors, who then republish results with slight changes while changing the title and rotating the names of first/last author. When the Journal of Prosthetic Dentistry began in 1951, the average number of authors for each article averaged about 1.5 names during the first decade, and increased gradually every decade, reaching 4+ in the 1990s. PubMed uses only four names.

*Posted by Philip A. Schwa..., 5/3/2013 7.00pm*

We routinely run submitted manuscripts through iThenticate software to detect issues related to plagiarism. Typically, the iThenticate report indicates a "similarity" level of less than 20%. We do have occasional papers that are in the 30-40% similarity range. The similarity is mostly

associated with duplication of methods regularly used by the authors' laboratory. We recently received a manuscript from a reputable laboratory (indeed, from a lab that publishes regularly in our Journal) that came through iThenticate with a similarity score of 61%. When we went through the report to understand the details, it was clear that the problem was again due primarily to repetition of methods – i.e., “self-duplication.” While we have no desire to force authors to “reinvent the wheel” every time they write a paper, we are concerned that authors are – more and more – establishing a template for their reports so that they can simply “plug in” the latest variable. In this case, the authors are reporting the effects of a new drug, using measures that the lab uses – and has published on – routinely. They had previously reported (in our Journal) on a different drug – but using the same procedures. What is the proper balance between asking the authors to provide “new” text for their manuscripts and not imposing an unnecessary burden on them?

*Posted by Dave Fernig, 11/3/2013 2.24pm*

New text versus old is an issue, but one can paraphrase (it is a skill and potentially more difficult for non-native English speakers/authors). Beyond that surely if one is doing something new, then the “angle” of the introduction should change over the years? However, in some fields gathering observations on a particular system over time across many papers is an essential exercise. Perhaps in such instances it is the journals that are at fault: what would be wrong with a short one paragraph introduction? This would help. Indeed most of the PhD theses coming out from my lab now have a short ~10 page introduction, rather than the UK norm of 35-50 pages and no one seems to suffer. Indeed, this has been welcomed by many (but not all) of the external examiners. Highlighted by others here is the question of data re-use. I have recent experience of bringing issues of data re-use to the attention of various journals (openly). It is clear that the problem is dealt with piecemeal. It would seem that serial offenders may end up with serial corrections from different journals, each obtaining retrospective copyright transfer from the original sources. However, the problem of serial data re-use is not dealt with and the offenders may continue. So some system is required to put a barrier in place for serial offenders. Retraction seems to be the best deterrent. This, after all is the medicine the institutions who employ many authors, Universities, prescribe to their students and it would be perverse if there was one law for students and another for their teachers.

*Posted by Sam Matthews, 11/3/2013 2.37pm*

I'd be interested to know what the justification might be for demanding a new and original introduction every time. If a lab is pursuing a drawn-out project, it's likely to publish several papers on the topic before it's through (at least in chemistry, my field). The justification for pursuing the research isn't likely to change over the course of the project, and the state of the art may not change significantly either (with the exception of the papers previously published, which ought to be cited, obviously). Given that, why should an author bother finding new ways to explain why their research has merit every time they submit a new paper? Why not develop a cogent, concise explanation of the relevant background, and use it every time, updating as necessary?



*Posted by Mark, 11/3/2013 10.44pm*

How about educating people about copyright and ethical practices in publishing? One also must take note of the fact not everybody (different cultures) share the same notion on plagiarism (self or otherwise)- the western concept of "owning the words". . My experience with students from different cultures (Asian, East Asian) reveal that most of them are appalled at accusation of plagiarism or duplicate publication. They don't seem to do it intentionally. Many research papers have addressed this issue. This Forum should take a more holistic approach of educating people. I strongly feel that self-plagiarism is misnomer or a oxymoron term. Better to say copyright violation.

*Posted by Soumitra Kumar ..., 12/3/2013 8.31am*

Regarding the retraction conditions I think it is somewhat terminating to consider joint products or by products in research and publications. So if the same process or almost identical process could lead to another conclusion why should that not be publishable as a separate paper by the publishers. It is the scientific contribution and the new knowledge that is created which is of essence. The reason I am saying this is because as many of us know perhaps all of us involved in research know and especially those coming from non English mother tongue countries know very well that to write a paper involving appropriate language takes a long time and effort which may be avoided if we follow a system of modular transformation that is using existing papers word processed as a process which can be changed in certain modules to give rise to another research output, this works for manufacturing processes, I don't see what is wrong with applying such an approach to research output, in fact I think publishers and the academic community especially Repec should encourage such a process.

*Posted by Alan Schwartz, 20/3/2013 5.43pm*

I strongly agree with two points previously made:

1. Text recycling is a much better term than self-plagiarism, which is an oxymoron. In addition, it's great to focus the policy on text recycling per se as distinct from copyright violation.
2. Richard Saitz's statement, which I reproduce here: "The guideline might state when text *\*should\** be recycled. Text should be recycled when a paper is reporting different findings from a study previously reported. Why should synonyms be used for methods? If the methods were identical they should be described using the same words. This might apply to papers reporting subgroup analyses from clinical trials or observational analyses of trial data. Similarly, when describing participant characteristics in those cases, hopefully they are the same, if the sample was the same. Using synonyms and mixing up words serves no purpose in those cases. On the other hand, there is clearly duplication that should not happen, like reporting the same discussion, or analytic results. Introductions could present similar arguments but would be unlikely to be justifiable as identical."