

Quality Control in RePEc ... why it is so hard?

Thomas Krichel
RANEPA & Open Library Society

Moscow 2017-04-20

Контроль качества для RePEc: предстоящие задачи

Томас Крихель
РАНХиГС и Общество открытой
библиотеки

Москва 2017-04-20

Thanks

- RANEPa for sponsoring my attendance.
- Organizers for tolerating me.
- I tend to be a bit obscure.

structure

- A bit about RePEc.
- To control quality, we need to sanction contents that is in RePEc.
- Assuming we want to impose a sanction on RePEc contents how can be do it.
- Why should we sanction anything?

RePEc and me

- Basically I am the founder of RePEc.
- In the sense that I wrote the documents that allow RePEc to actually work.
- In the sense that laid all the ground work years before RePEc was created.
-

working papers

- A publication from that is specific to economics and computer science.
- It remains active in economics only.
- Economics departments issues early accounts research papers written by their staff.
- In the early 90s, they were still circulated by postal mail.

1991

- Contents: Warwick working paper acquisitions list.
- Technology: Email lists.
- Idea: distribute the acquisitions list through email.
- leads to the formation of BibEc.

1992

- Contents: free software for TeX, emacs etc.
- Technology: anonymous ftp.
- Idea: make academic papers available in a public place on the Internet for download.
- Leads to the formation of WoPEc.

NetEc

- It was a group of Internet services to help economists.
- I first ran it in February 1993 on a gopher server that Manchester computing centre gave me.
- On the web since 1994.
- Essentially run by me.

BibEc 1993 → 1997

- Based mainly on acquisitions data for printed economics working papers from the Documentation Centre of the Economics department at the University of Montreal.
- Run on a volunteer basis by Thomas Krichel and Fethy Mili
- Holdings go back to the late 1980s, around 40,000 items
- What quality control?

WoPEc 1993 → 1997

- Central collection of bibliographic data on electronic, network accessible working papers.
- Initially unpaid volunteer work by Jose-Manuel Barrueco Cruz and Thomas Krichel
- Between 1996 and 1998 JISC funding allowed to hire Jose Manuel full time.
- Reaches 5000 papers in 1997.
- What quality control?

EconWPA

- This was an independent effort by Robert P. Parks to build a centralized working paper collection.
- The idea was to import the technology that Paul Ginsparg had developed for arXiv to build a central collection of working papers.
- This makes for the possibility to centrally build quality control.

WoPEc vs EconWPA

- It's the difference between a preprint system and a working papers system.
- Imposing a preprint culture on Economics did not work.
- But in the mean time there had been a lot of discussions that produced a lot of heat but not much lights.

William L. Goffe 1995

- “What I would suggest is this: a distributed system with any number of sites, each mirroring each other. [...] archives could “join” the system (say it was written in perl so could run on NT as well as Unix). Then you'd have the best of both worlds [...] Such a system could easily grow with the profession's use of the net. Such a system would GREATLY benefit the profession.”

RePEc

- RePEc is essentially a way to implement the vision of Bill Goffe.
- I set to work out a protocols to implement the vision.
- But we note neither Bill nor Thomas did not have quality control in mind.

RePEc principle

- Many archives
 - archives offer metadata about digital objects (mainly working papers)
- One database
 - The data from all archives forms one single logical database despite the fact that it is held on different servers.
- Many services
 - users can access the data through many interfaces.
 - providers of archives offer their data to all interfaces at the same time.

what is in an archive?

- It hold two principal (own term here) type of records
 - An archive record
 - A set of series records
- It may hold document records.
- This is best understood by examples.

sample archive data

Template-type: ReDIF-Archive 1.0

Handle: RePEc:sur

Name: School of Economics, University of Surrey

Maintainer-Email: i.lazopoulos@surrey.ac.uk

Description: This archive provides research papers from the School of Economics of the University of Surrey, in the U.K.

URL: <https://repec.som.surrey.ac.uk/RePEc/sur/>

series data sample

Template-type: ReDIF-Series 1.0

Name: ROME Working Papers

Provider-Name: ROME Network

Provider-Homepage: <http://www.rome-net.org>

Maintainer-Name: Albrecht F. Michler

Maintainer-Email: michler@hhu.de

Type: ReDIF-Paper

Handle: RePEc:rmn:wpaper

archive vs series

- The archive is a place where data is held that flows into RePEc.
- Once you have an archive, you can open as many series as you like.
- What are the consequences for quality control?

sanctions

- Let's us a typical economics approach.
- Assuming RePEc want to sanction some contents ... for what reason.
- How could this sanction be implemented.

sanction classes

- There can be two classes of sanctions.
- There are subtractive sanctions.
- There are additive sanction.

additive sanctions

- RePEc--or same agent or agents--could post records with
 - list of sanction record
 - for each sanctioned record the type of sanction
- We could create a controlled vocabulary of sanctions.

problems with additive sanction

- A formal specification of the format that the addition takes does not exist and may be hard to do.
- RePEc has no way impose implementation of any additive sanction, by “itself” or by an agent.
- We don't need to do it. Anybody can.
 - If the sanctions issued use RePEc handles, I don't see any problem to provide an agent within the host in the RePEc domain and they can operate a sanction site.

subtractive sanctions

- Since archive maintainers may add any series, and any documents in the series, the single subtractive action that we can take is the removal of an entire archive.
- Hang on ... maybe we can't even do that.

core archive

- RePEc management (mainly Kit Baum) maintains a server with the archive and series templates of all RePEc archives known to RePEc.
- We can withdraw the description of an archive on the core archive.
- The main RePEc services will delete all copies of the template and presumably of the data.

archive withdrawal problem

- It is a very crude tool.
- The data may be copied somewhere else.
- There are no policies on any record withdrawals in RePEc.
- Example, for the OAI-PMH gateway I now delete records after three years.

not open ...

- We could not allow a publisher that looks suspicious to open a RePEc archive.
- This could be done
 - archive code issue
 - when archive is registered in the core archive

filter at archive code issue time

- Organizations that wish to open an archive have to get a code that grandfathers other identifiers in the archive.
- Many organizations ask for a code, but don't finish the archive. It's quite a lot of work as the protocols are not standard.
- Filtering at this stage seem to place a big burden on Kit or whomever would assist.

filter at archive listing time

- Well, at that stage the provider has done work to build the archive.
- It would be unfair to not allow the provider access the RePEc network.
- A pre-emptive warning that the contents will later be reviewed will reduce incentives to
- and it's not clear who will do the job.

why do we need to do anything

- Well last year at this meeting Ольга Владимировна Кириллова told me that
 - there are a some publishers listed in RePEc
 - produce junk contents
 - they use the listing to promote
- I later asked her for a list but I don't have one at this.

first step taken

- I wrote a blog post
- <https://blog.repec.org/author/krichel/>
- The emphasis was on quality control of records going forward.

from the blog post (1)

- There is a growing number of journal-like outlets that pretend to be normal open access journals. But in reality, all they do is take authors' money, and put the content up on a web site. They do no quality control. They have no editorial board that does any work. In fact, many times people on the board do not even know that they are on it.

from the blog post (2)

- Traditionally, RePEc has not done any quality control prior to listing additional journals. We believe that quality can best be assessed by users of the RePEc dataset. However, we have been criticized for helping these deceitful outlets gain a mantle of respectability through their RePEc listing. Therefore we take this step forward. We expect quality control also to be an issue with toll-gated journals.

from the blog post (3)

- The volunteer we are looking for will determine the exact name of the committee and its remit. (S)he would recruit a few committee members. (S)he would run the mailing list and maintain some web pages for the committee. RePEc can provide both. Anybody who is interested in this work should contact repec@repec.org.

comments on the post

- It's the work on a genius.
- But it is misleading to talk about “quality control”.
- A lot of people (mis)took this as look at the fundamental quality of journal.
- The essence of the post pointed at a fraudulent activity that needs to be found out about.

reactions

- Well about 10 reactions to the blog post itself from volunteers.
- Then about 20 more when Christian Zimmermann placed the text into his monthly mailing to RePEc authors.
- We needed a mailing list to get them to talk to each other.

chicken and egg

- To open a mailing list, we need some sort of a name for the list.
- It should be related to the work on the committee.
- But that means a name of the committee needs to be found.
- I settled on the name “committee on deceptive publishing”.

deception

- We know there are good papers and that there are lousy papers.
- We know that there are good journal and that are
- We can not draw an objective line separating them.
- We can show, maybe even proof deception.

state of work

- There has been discussion on the name.
- Somebody suggested “task force”
- I recognized the need for a written constitution.
- I feel that among the people in the group, we would have problems finding somebody who understands.
- The discussions are public. <http://lists.openlib.org/pipermail/codep-run/>

ad hoc case Bentham

- When Bentham applied for a code, I took the initiative to put it in front of the committee even though we don't have a constitution.
- They only have one journal that—with a stretch—is relevant.
- We found no evidence of deception.
- I regretted this step.

2017-01 a pioneering additive sanction

- Christian Zimmermann wrote
- “Following up on the recent conversation about journals not in the scope of RePEc: I have now removed some from the listing on IDEAS. Their content is still indexed and available, the journals are just not listed.
- A listing of those not listed is here:
- <https://ideas.repec.org/i/hidden.html>”

Спасибо за внимание!

Томас Крихель

<http://openlib.org/home/krichel>