Review of: "Reflections on the Misuses of ORCID iDs"

Silvio Peroni[1]

1 University of Bologna

Metadata

Title: Reflections on the Misuses of ORCID iDs

Author: Miriam Baglioni, Andrea Mannocci, Paolo Manghi, Claudio Atzori, Alessia Bardi, Sandro La Bruzzo

Submitted to: 17th Italian Research Conference on Digital Libraries (IRCDL 2021)

Review

In this article, the authors reflect on existing and potential issues and errors in the use of ORCIDs. Their analysis is entirely derived from the work they have done to produce the OpenAIRE Research Graph, particularly during the deduplication phase of the production workflow.

The article is well-written and organically presents such issues, supporting the discussion with real data in the ORCID repository. It is important to spot these situations to devise appropriate strategies for addressing them in different contexts of use.

Only a few suggestions where there is room for improvements:

1. The authors say that "an analysis of the email address domains could enable ORCID to identify non-institutional emails and notify the authors for a change". However, this may present some issues when we consider independent researchers, i.e. scholars without an affiliation who expose their personal email address. Even if they are rare, I have seen some examples of such situations, particularly in the humanities domain. Thus, it would also be crucial to see if a non-institutional email address refers to a real independent scholar or not. I believe a careful check of this situation should be done by hand somehow since one could always lie to that respect. However, this would need much manual curation which, I presume, is very difficult to address - even if this depends on how much non-institutional email addresses are in place in the ORCID system. Could it be possible for the authors to reflect on this situation in the discussion?

2. The authors presented a systematic overview of existing errors that may derive from the misuses of ORCIDs. However, it is not clear if they also know the proportion of these misuses in the ORCID repository. In case they know them, it would be great to report these numbers in the article to show explicitly the impact that each issue has overall. Otherwise, it would be good to add some reflection about possible future works in which the authors could come out

with quantitative measurements of each of these issues.

3. Footnote 4 is useless - I would suggest citing [8] instead.

Summarising, this article deserves to be presented in IRCDL 2021, since it addresses a significant problem of interest for the community.