



HAL
open science

RePEc Author Service: An established community-driven PID

Pablo de Castro, Ulrich Herb, Laura Rothfritz, Joachim Schöpfel

► **To cite this version:**

Pablo de Castro, Ulrich Herb, Laura Rothfritz, Joachim Schöpfel. RePEc Author Service: An established community-driven PID. Knowledge Exchange. 2023. hal-04245155

HAL Id: hal-04245155

<https://hal.univ-lille.fr/hal-04245155>

Submitted on 16 Oct 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

RePEc Author Service: An established community-driven PID

February 2023



This case study is part of a series that has been produced within the study on “Risks and Trust in pursuit of a well-functioning PID infrastructure for research” commissioned by the Knowledge Exchange in July 2021. The main outcome of this study is a report examining the current PID landscape with an emphasis on its risks and trust-related issues.

This complementary series of case studies aims to provide a deeper insight into specific areas of activity, workflows and stakeholders within this wider PID landscape.

Title: RePEc Author Service: An established community-driven PID

Authored by: Pablo de Castro, Ulrich Herb, Laura Rothfritz and Joachim Schöpfel
(scidecode science consulting).

Email: office@knowledge-exchange.info

DOI: [10.5281/zenodo.7330516](https://doi.org/10.5281/zenodo.7330516)

All content published is shared under a Creative Commons Attribution licence
(CC BY 4.0) creativecommons.org/licenses/by/4.0



Contents

1. Rationale	4
2. RePEc and its Author service (RAS)	6
2.1 RAS and the RePEc services	6
2.2 Uptake	8
2.3 Sustainability & Governance	8
2.4 Funding parties	9
3. Issues around risks and trust	10
3.1 RePEc as an (un-)organization	10
3.2 Funding & Sustainability	10
3.3 Cross-functional integration vs. functional self-restraint	11
3.4 Evidence of functionality and community needs	11
3.5 Open Data	11
3.6 Summing up	12
4. Authorship	13

1. Rationale

The RePEc Author Service (RAS) is a useful addition to the DAI case study, as it has similarities to DAI in that it is a non-profit, community-based service, but also differences in that it is disciplinary, and especially as it survived the advent of ORCID.

RAS is notable for its connection with a variety of other services (document download statistics, bibliometric analysis, citation counts, person and institution rankings) that reside within a kind of RePEc service family.¹ Strikingly, RePEc or the RAS operates its own affiliation manager that identifies institutions down to the department level² and neither implemented ORCID and ROR nor synchronizes data with these services. Selecting RAS as a case study provides valuable information about why RAS (despite the competing ORCID and ROR initiatives) still exists (while, for example, the DAI no longer does) and how this relates to community support and funding. A first insight is that RAS, perhaps more than DAI, which may have been a more organisation-driven effort, is actually being adopted by researchers themselves and by that is (just as the PIDs

for research instruments and facilities or IGSN) a technical PID. Possibly its disciplinary nature also plays a role for its resilience, contributing more to the profiling of institutions and individuals (also in the linked RePEc cosmos). At the time when the DAI was operational, there was no application for this use case with DAI, whereas it is evident in ORCID through integration with databases such as Web of Science or Scopus. Also of interest is the question of funding (also to be retrospectively considered): currently RePEc is largely supported by the Federal Reserve Bank of St. Louis.

1. <http://repec.org/>

2. <https://ideas.repec.org/top/top.econdept.html>



2. RePEc and its Author service (RAS)

2.1 RAS and the RePEc services

Research Papers in Economics (RePEc) emerged from the WoPEc (Working Papers in Economics) project³ funded between 1996-1999 by the Joint Information Systems Committee (JISC). It was launched in June 1997 to decentralize organizationally the work done by WoPEc and thus make it independent of grant needs. It is one of the world's largest databases for Economics, with more than 1.3 million titles. RePEc with its various services claims to have become a main entry point for searching working papers, journal articles and conference proceedings. The listing of a series of papers or a journal in RePEc offers many advantages, both for the individual papers or articles and for the series or journal as a whole:

- ▶ direct dissemination of papers to the community via the established, freely accessible RePEc services IDEAS and EconPapers, as well as via the alerting service New Economic Papers (NEP)
- ▶ detailed download statistics via LogEc
- ▶ detailed citation analyses and integration into CitEc
- ▶ integration of titles into the author profiles of the RePEc Author Services (RAS)

RePEc promotes⁴ the advantages of its Author Services with these features and services making use of RAS: author identification/ disambiguation/ profile, notifications, OpenID credentials and author and institution rankings. RePEc also emphasizes that RAS was already launched

back in 1999 to indicate its reliability. Not unlike ORCID it relies on a self-registration, this registration triggers the creation of a RePEc short-ID, a unique and permanent code which is used in RePEc services as well as external services as WikiData⁵ to uniquely identify authors. After the registration, authors may claim their works (supported by suggestions made by RePEc), subsequently their profile (including name, affiliation and works) is set up. These profiles are also used in other RePEc services mentioned above.

Registered authors are also notified about newly found citations and other statistics about the visibility of their works in form of downloads. Users may also follow authors to receive updates on their publications. As RAS' authentication is based on the OpenID protocol, the credentials can be used for other RePEc services as well. The RAS information is combined with other data from other RePEc services to produce several rankings or other bibliometric services (co-authorship networks, CollEc project⁶) and other tools as a kind of academic genealogy tree for economics (RePEc Genealogy⁷).

During the registration process, authors are required to choose an affiliation from RePEc's affiliation manager EDIRC (Economics Departments, Institutes and Research Centers) or to suggest a new one. Newly suggested affiliations are obviously added automatically to the service (and also removed after some time, if no publications are linked to them), the affiliation index is not based or synchronized with ROR, also there is no integration of ORCID into the RAS - authors may not even add an ORCID to their RAS profile. EDIRC covers according to its website currently 15,076 institutions in 232 countries and territories.⁸

³ <https://www.ukoln.ac.uk/services/elib/projects/wopec/> WoPEc in turn was part of NetEC which "was a collection of projects that aim[ed] to improve the scholarly communication in Economics" <http://openlib.org/home/krichel/hisn.html>

⁴ <https://blog.repec.org/2019/03/27/why-authors-should-have-an-account-with-repec/>

⁵ <https://www.wikidata.org/wiki/Property:P2428>

⁶ <https://www.wikidata.org/wiki/Property:P2428>

⁷ <https://www.wikidata.org/wiki/Property:P2428>

According to an email conversation with RePEc, the service considers including features that allow adding and import other identifiers, perhaps WikiData, for starters.

The utilization of the RAS profile within other RePEc services may be illustrated especially by its integration in IDEAS, which displays the RAS information⁹ in the IDEAS profile. It comprises the basic personal information, information about affiliations (which may be split proportionally between several organizations), publications, citations to these (both from scientific publications and blogs), download statistics, a researcher's position in an abundant number of rankings, co-authorship networks (providing network analysis measures as closeness, betweenness, centrality, etc.).¹⁰ RAS is also used within the CitEc

service to generate more detailed citation data and bibliometric information (as e.g. co-authorships or recent citations)¹¹:

Andrei Shleifer : Citation Profile

98
H index

188
i10 index

66016
Citations

RESEARCH PRODUCTION:

174 Articles	375 Papers	5 Books	11 Chapters
------------------------	----------------------	-------------------	-----------------------

EDITOR:

2
Series edited

RESEARCH ACTIVITY:

- 37 years (1984 - 2021). [See details.](#)
- Cites by year: 1784
- Journals where Andrei Shleifer has often published
- Relations with other researchers
- Recent citing documents: 4699. Total self citations: 243 (0.37 %)

MORE DETAILS IN:

[EconPapers](#) [Author's homepage](#)

ABOUT THIS REPORT:

- Permalink: <http://citec.repec.org/psh93>
- Updated: 2022-02-28 RAS profile: 2022-01-18
- Missing citations? [Add them](#) Incorrect content? [Let us know](#)

8. <https://edirc.repec.org/>

9. In the following, Andrei Shleifer's profile is used for illustrative purposes, <https://ideas.repec.org/e/psh93.html>

10. https://app.collec.repec.org/app_direct/collec_app/

11. <http://citec.repec.org/p/s/psh93.html>

2.2 Uptake

While ORCID has (according to its website¹² with date of 20th September 2022) 14,959,053 registered IDs, RAS indicates (as of 20th September 2022) 65,169 registered authors (which have authored 1,878,499 items listed in RePEc)¹³.

RePEc states that its data is re-used e.g. in the databases EconLit, EconStor, OpenAIRE, Sciverse and EBSCO databases. Nevertheless, it is unclear to what extent these very rich data and valuable services of RePEc are integrated in external databases or services. Most databases mentioned by RePEc as re-using its information are probably simply importing its bibliographic data. Also, the Leibniz Information Centre for Economics (Leibniz-Informationzentrum Wirtschaft ZBW) does not reuse RePEc PIDs (RAS and EDIRC) in its own services, but citation data from CitEc. Nevertheless, in 2017 the ZBW used Wikidata to match around 3,000 authors with RAS IDs and the Integrated Authority File IDs (GND) by their publications¹⁴ to allow a more accurate attribution of publications to authors in their EconBiz¹⁵ portal.

2.3 Sustainability & Governance

Shortly after the acquisition of SSRN by Elsevier, RePEc published a statement¹⁶ pointing out that it would be “impossible” to buy RePEc. The key argument says that RePEc is not attractive for commercial stakeholders, as it (even though generating values for the community it serves) has no financial value. Operating the service without secure funding is only possible as RePEc is

optimized to run at extremely low cost, so its services are offered for free. This in fact requires volunteer work¹⁷ and sponsorship for hardware, hosting and bandwidth. As all RePEc data is openly available (even in public domain¹⁸) there is, in RePEc’s opinion, no chance of selling its data to generate noticeable revenues. In fact, this open data approach would allow to re-build in the case of RePEc’s failure its services freely from scratch. This point was also made several times in expert interviews during this study. The fact that RePEc appears to have very low operating costs is an indication of resilience and sustainability, which is only found in other services at a higher investment.

Considering RePEc’s launch already back in 1997 this approach so far did not prove to be unsustainable. Sticking to it, RePEc is asking its community for support in a range of areas as for instance RePEc archive maintainer (curation of publications submitted by an institution), NEP report editor (dissemination of new online working papers via weekly reports), editor for the Munich Personal RePEc Archive (a kind of orphan repository), help with EDIRC, RAS, RePEc Genealogy, RePEc Biblio, RePEc’s Plagiarism Committee. RePEc also asks to contribute hardware or hosting services. Regarding EDIRC RePEc mentions:¹⁹ “This database has grown tremendously, and help is needed to maintain it.” This sheds light on the administrative and curation burden when operating an affiliation manager. The support needed with RAS is described as help “in monitoring the activity on the service and managing the profiles of deceased authors.”

¹² <https://orcid.org/statistics>

¹³ <https://ideas.repec.org/stats.html>

¹⁴ <https://zbw.eu/labs/en/blog/wikidata-as-authority-linking-hub-connecting-repec-and-gnd-researcher-identifiers>

¹⁵ <http://www.econbiz.de/>

¹⁶ <https://blog.repec.org/2016/05/17/repec-is-independent-and-cannot-be-bought/>

¹⁷ As of March 29th, RePEc mentions twelve team members, <https://ideas.repec.org/team.html>

¹⁸ A corresponding statement can be found here: <http://repec.org>

¹⁹ <https://ideas.repec.org/volunteers.html>



2.4 Funding parties

Information about the funding of RePEc are sparse and scattered around its website, this mentions in different sections the Federal Reserve Bank of St. Louis, which offers hosting capacities since 2011²⁰. This organization is also the employer of Christian Zimmermann, who is obviously one of the main contact persons for RePEc, why it may be considered RePEc's most important sponsor. In an email conversation, Christian Zimmermann also stated that he would be allowed to spend some of his working time on RePEc.

Additionally, in the context of the RePEc service ArchEC that provides long-term archiving of RePEc templates and full-text files there are three sponsors mentioned²¹: a) the Fondation Banque de France, b) "an important economics research organization based in the United States, who pay for the server" (probably the Federal

Reserve Bank of St. Louis), c) the Open Library Society in New York. The last one mentions on its website²² one of RePEc's establishers back in 1992²³ as its founder, Thomas Krichel. This makes it seem possible that the Open Library Society was set up to organize external funding through third parties for RePEc.

Not only RePEc as such is hosted by a sponsor, the same is true for at least one of its services, the co-authorship analysis tool COLLEC,²⁴ that is according to its website hosted by Symplectic. Symplectic is one of the services offered by Digital Science, a commercial provider in research analysis. Also, EconPapers, another RePEc service, relies on funding, in this case by the Örebro University School of Business.²⁵

²⁰. <https://ideas.repec.org/history.htm>

²¹. <http://archec.repec.org/>

²². <http://society.openlib.org/>

²³. <https://ideas.repec.org/history.html>

²⁴. https://collec.repec.org/app/collec_app

²⁵. <https://econpapers.repec.org/>

3. Issues around risks and trust

3.1 RePEc as an (un-)organization

RePEc's organizational form seems quite undefined. It is also not clear (at least from an outside perspective) how the community is engaged in plans for further development of RePEc and RAS.

There are two mailing lists available, RePEc-announce (with general information about RePEc developments) and RePEc-run (with discussions of technical matters regarding RePEc services), but the overwhelming number of mails on the first one are postings contributed by two RePEc team members while access to second list's archive is restricted. One feedback channel may be the volunteers supporting RePEc's operation through their manpower - even if there is no indication to what extent the volunteers have any influence on any strategic or organizational decisions.

The degree to which RePEc and its PIDs (RAS and EDIRC) depend on volunteers may reflect a sort of weakness (risk) and strength (trust) at the same time: Having a larger volume of continuous funding or another more formalized governance structure might weaken the responsibility the community feels for RePEc. On the other hand, this kind of loose organizational structure requires a lot of trust in the service and the people behind it for authors and institutions to make use of it. Mentioning this, however, the age structure of the RePEc makers and team members²⁶ may be seen as critical for a service obviously relying so much on individuals, their knowledge and expertise. According to an email conversation with RePEc both issues of technical viability through switching to up-to-date web/ database techniques and team members succession plans are to be addressed.

3.2 Funding & Sustainability

Apparently RePEc organizes its funding

- ▶ through the Open Library Society (instead of e.g. SCOSS) which was launched by one of the inventors of RePEc which may indicate a strong emphasis of the service's independence.
- ▶ within its own organizational network, what can - if successful - of course be considered in indicator of trust:

“

So one important measure for me whether they trust in the system is whether they invest in it. Whether they put resources to it. This is in particular, with government agencies, when they make a commitment. Once they make this a line item, they stick with it. But to get to that point is hard.

”

One reason for this mistrust in centralized approaches to organize or finance infrastructures may lay in the fact that ORCID used RAS data as seed data to populate its databases, but deleted the RAS identifiers afterwards (as reported by Christian Zimmermann), which probably provoked the feeling of having been exploited.

Focussing on the sustainability, RePEc seems either to rely heavily on external funding without revealing much information about that or really manages to operate quite successful with its minimum cost approach. Having a look at the huge number of services that were built around RePEc, some surely are of great importance and have

²⁶. <https://ideas.repec.org/team.html>

significant value, while others perhaps may be results of some sort of “play with the data” attitude. This does not necessarily indicate a risk, it may even reflect that RePEc even is operating his business so balanced that they can even afford to set up services without any specific demand for an evident use case.

3.3 Cross-functional integration vs. functional self-restraint

The fact that the PIDs operated by RePEc (RAS, EDIRC) seem not be integrated in external services nor seem to offer any added values such as automated metadata retrieval may be considered either a risk (as this may reduce its acceptance in an ecosystem of increasing demands on interoperability), but may also be an indicator that RePEc is a solid system in itself with solely disciplinary use cases and therefore may not need interoperable functionalities.

3.4 Evidence of functionality and community needs

At first, operating a service as RePEc since 1997 (and RAS since 1999) may be considered a very strong trust indicator, the more as it was obviously able to cover its operational expenditures even long before invests in open infrastructure were broadly discussed.

“

Because well, okay, there are a few issues here and there, but it's been running for a very, very long time already.

”

The service itself is still considered to be highly significant in Economics, especially because of the researcher and institution rankings, especially of non-university economic research, and services as citation counts. For example, the ifo Institut – Leibniz-Institut für Wirtschaftsforschung an der Universität München e. V. uses RePEc data intensively to produce its rankings.²⁷

3.5 Open Data

RePEc's own argument in its trustworthiness, that data in public domains allows anyone to set up the service again for free (a principle also put forward by interviewees), is true, but assumes that there are documented database dumps available, which might not be taken for granted.

“

And so the... to my mind, the thing that an organization that's running infrastructure, or any service can do, that's most useful and the best insurance that they can't be coopted, is to make the data as forkable as possible. Now, clearly, the data and the service, not just not just the data. So that, you know, if a sufficiently large part of the community is unhappy with what the organization is doing, they can go and start another way. Right? And there are no artificial, technical, or data barriers to them doing that. Clearly, there's always the barrier of bringing the rest of the community with you. But if you can't do that, then the question is, are you actually an outlier? I mean, so it's sort of a self reinforcing thing.

”

²⁷. <https://www.ifo.de/en/node/21206>

“

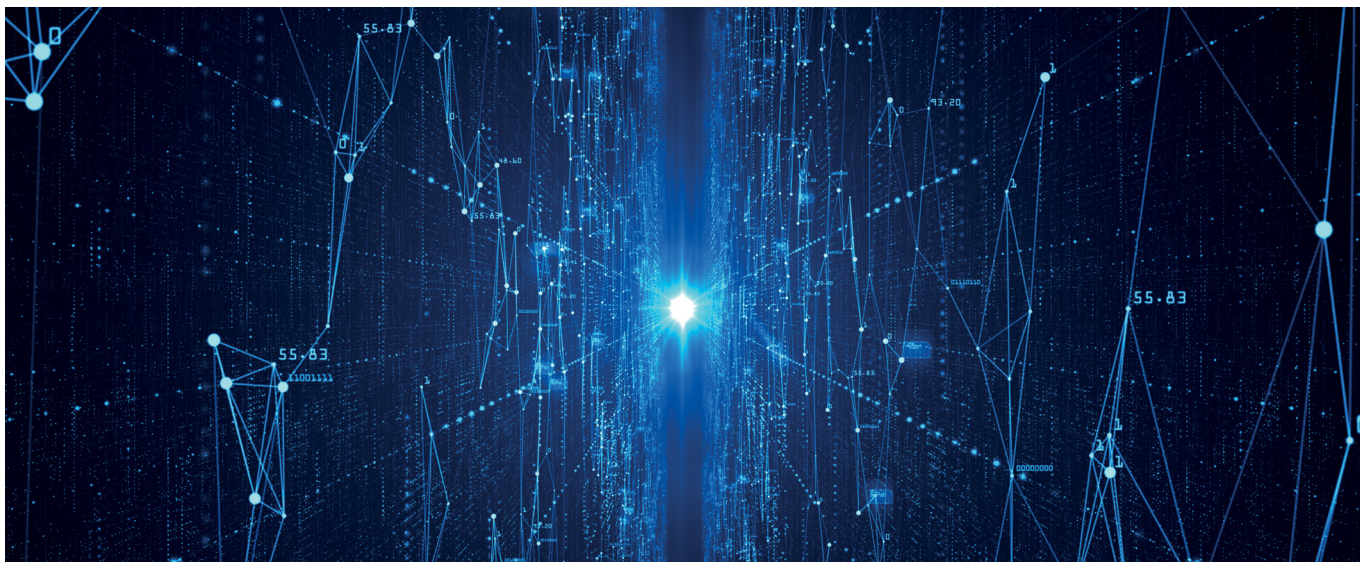
And that's really important that organizations and infrastructures are addressing these and so, as an example, at DataCite, making our data far available and the CC0 and making it public deposits or archive of that and a data dump available is important, having open API is having the code Open Source, these are important insurance mechanisms to make sure that the community has that assurance that it would be long term that these things would still be in place, that technology and the data would not disappear.

”

3.6 Summing up

Contrasting RAS (RePEc and EDIRC) against e.g. the Digital Author Identifier (DAI) or ORCID (or ROR) the RePEc services seem to benefit from serving the needs of a community of scientists and organizations, whereas ORCID and ROR serve - at least in Economics - more the needs of libraries, universities, funding agencies and publishers, so co-existence of RAS/ EDIRC and ORCID/ ROR seems possible, unlike DAI which addressed the same needs as ORCID. RePEc seems to be a service

whose disappearance would pose very serious problems for the Economics community, since it offers functions that are highly valued and widely used in the community. RePEc never gave up the sovereignty about technical, organizational and strategic decisions - what probably was only possible as a result of the trust in it that stimulated financial and human resources support by organizations that benefit from RePEc. Just as with IGSN (see the respective case study) the benefits the service offers strongly derives from (and depends on) the incorporated knowledge of some of the individuals who build/run the service. However, while IGSN seeks solidity by following best practices, largely generic solutions, and seeking strategic partnerships, RePEc appears to be a one-of-a-kind technical/organizational entity whose persistence depends very much on individual commitment and knowledge - success factors that may be difficult to transfer to others. Hence, the issue arises of how to keep RAS operational if, for example, it is not possible to transfer or generalize the expertise concentrated in a few. This raises the concern whether it would be possible to migrate RAS to other PIDs such as ORCID in order not to lose its functionalities altogether. This example may be hypothetical in the RAS case, since the discontinuation of RAS would presumably be accompanied by that of RePEc - however, this poses the general question of procedures for the orderly replacement of one PID system by another.



4. Authorship

This case study has mainly been written by Ulrich Herb (Saarland University, ORCID <https://orcid.org/0000-0002-3500-3119>) within a team of consultants including Pablo de Castro (University of Strathclyde and euroCRIS, ORCID <https://orcid.org/0000-0001-6300-1033>), Laura Rothfritz (Humboldt University Berlin, ORCID <https://orcid.org/0000-0001-7525-0635>) and Joachim Schöpfel (University of Lille and euroCRIS, ORCID <https://orcid.org/0000-0002-4000-807X>) under the umbrella of scidecode science consulting (ROR <https://ror.org/02c0bjd31>). The work has been overseen by the Knowledge Exchange Task & Finish Group whose composition is listed at <https://www.knowledge-exchange.info/event/pids-risk-and-trust>.

Knowledge Exchange
C/O Jisc
4 Portwall Lane,
Bristol, BS1 6NB
United Kingdom

T +44 203 697 5804
E office@knowledge-exchange.info