



13th International Conference on Current Research Information Systems, CRIS2016,
9-11 June 2016, Scotland, UK

SocioRePEc CRIS with an interactive mode of the research outputs usage

Sergey Parinov*

Central Economics and Mathematics Institute of RAS, Nakhimovsky 47, Moscow 117418, Russia

Abstract

This paper discusses an approach allowing researchers – users of some CRIS – to initiate and maintain scholarly pre-publication communication. The pre-publication communication takes place between authors of some research outputs and researchers who use these outputs on a stage of preparing their own research papers. Such communication is initiated when a researcher who uses CRIS tools is manipulating with a research output text and/or with the text fragments to create his/her own research outputs. In this paper a SocioRePEc CRIS is presented, which is being designed specifically to allow researchers in economics (who belong to a bigger international community of RePEc users) to make experiments with pre-publication communication. By this, researchers get capacity to use research outputs available at RePEc (metadata and/or full texts in PDF) in so-called “interactive” mode. The discussion explains that this interactive usage mode leads to better adjustment of the researchers’ supply and demand, and opens an opportunity for the further wider/deeper research cooperation.

© 2016 The Authors. Published by Elsevier B.V.
Peer-review under responsibility of the Organizing Committee of CRIS2016.

Keywords: SocioRePEc CRIS, pre-publication scholarly communication, research outputs usage, RePEc dataset and users

* Corresponding author. Tel.: +7-916-914-9051.
E-mail address: sparinov@cemi.rssi.ru

1. Introduction

In the CRIS-2010 conference paper [1] some challenges for a CRIS-CERIF development were discussed in relation with the SocioNet project:

- How should we construct in form and function a system for shaping and sharing research outputs/results of individual researchers stored at institutional repositories (IR) or/and CRISs so as to provide efficient scientific circulation and necessary conditions for its maximal usage? [1]
- How should we organize a process of research outputs/results usage, and design necessary metrics and tools to provide maximally comprehensive and accurate statistics on the uptake, usage and impact of research results? [1]

Now, six years later, it has become clear that there are many successful approaches, concepts and available research information systems which demonstrate a real progress in meeting those challenges. Many researchers currently actively use available tools and services to discover, analyze and reuse research outputs, e.g. at Academia.edu, F1000, Google Scholar, Hyposes.is, Mendeley, ResearchGate, and many others (see the results of the Survey of scholarly communication tool usage in [2]). It is also important to mention a successful development of research outputs metadata aggregator-systems, e.g. OpenAIRE, DOAR and some others, based on OAI-PMH protocol, or using the RDF-LOD concept and the SPARQL protocol, like DBpedia.org, etc., which all improve conditions for research outputs circulation and reuse.

At the same time the Russian SocioNet team has been continuing its own research to meet the challenges listed above. The members of this team have designed some new forms of the research outputs usage by integrating the semantic linkage technique into CRIS functionality [3], [4]. As a result, a pilot of the open semantically enrichable research information system for researchers [5] has been provided, which allows authors of the papers to enrich their publications' metadata, including an ability to specify motivation of using the cited publications listed in the reference section of their papers. If authors of both – the citing and the cited papers – have their contact data (e-mails) at CRIS, the system can initiate some communication between them immediately after the author of a citing paper specify their motivation in the system: why, or how, or for what purposes he/she has used the cited paper.

After more than one year of experimenting with the different modes of scholarly communication, which can take place while one researcher is using in some way at CRIS the papers related with other researchers [6], it is possible to come to the conclusion that a very interesting specific case of scholarly communication springs up: the pre-publication communication between the author of a research output and a researcher who has used this output to create his/her own research output.

Unlike the already implemented case – when a researcher can enrich his/her papers' metadata after the papers have been published [5] - now a researcher can be given an ability to manipulate in some ways with research output text fragments at the very moment while (s)he is actually reading this paper at CRIS content. If some fragment of a paper is selected by a researcher, CRIS allows the researcher to make some actions with this fragment. According to an initial set of actions, a researcher can: a) use the fragment as a base for creating his/her micro research outputs for personal or public reuse; b) classify selected fragment and save it for further personal usage or share it with a group of colleagues or with the whole community; and c) make a private or public comment to the selected fragment (the same way as it can be made already in some systems, e.g. at Hypothes.is, ResearchGate, etc.).

As a result of such actions a researcher creates at CRIS the public information objects of two types: a research relationship or a research artifact. The research relationship expresses an opinion or a comment of a researcher about the content of the selected fragment. The research artifact is a micro research output created by a researcher for further reuse and it can present a new research idea, assertion, or other types of an intellectual product.

If such actions create public information objects, then a research information system can initiate direct communication between the authors of publications, from which the fragments were taken, and the researchers, who used these fragments to create new research relationships or artifacts. CRIS can send notifications to the authors immediately when it registers a fact of his/her publication fragments usage. The main difference from the already implemented case is in the fact that now communication will take place at the pre-publication stage, i.e. before a researcher composed his/her micro outputs as a manuscript and publish it as a traditional paper.

The idea to manipulate with fragments of publications is not new. In 2012 Neylon wrote: “So in short, publish fragments, comprehensively and rapidly. Weave those into a wider web of research communication, and from time to time put in the larger effort required to tell a more comprehensive story.” [7].

If researchers have an opportunity of the pre-publication communication they can work in an interactive mode of a research outputs usage. The interactive usage mode leads researchers to better adjustment of their research supply and demand and opens new opportunities (provides a challenge) for the wider/deeper research cooperation.

Thus interactive_research usage mode can become a real alternative or replacement to the “passive” mode provided by the traditional scholarly communication scheme based on the activity of academic publishers and journals.

In general, the basic idea behind this is a public visualization of the processes of the researcher’s scientific analysis and creativity which usually appears to be hidden from the research community as being carried out in the mind of a researcher. The traditional technique of the scientific citations in that case does not work well because it typically operates with the entire paper but not with the fragments containing the research results, and also it does not allow a researcher to explicitly indicate the motivation for the citations.

Neylon discussing technology to manipulate with publication fragments wrote: “This requires tools that are hard to build, standards that are hard to agree, and cultural change that at times seems like spitting into a hurricane. Progress is being made, in many places and in many ways, but how can we take this forward today?” [7].

The authors of the Publons research information system, which aims to generate innovative scholarly communication, wrote: “It is hard to say exactly what these new tools will be. Their development will require a great deal of experiment and iteration. Therefore, instead of attempting to a priori identify and build the right tool, we need to develop a framework (and culture) in which experimentation with different forms of collaboration is encouraged.” [8].

By this paper a presentation of a specific research information system, called SocioRePEc (<https://sociorepec.org/>) is initiated. This system is being designed for the experiments with an interactive mode of the research outputs usage based on pre-publication scholarly communication.

Since SocioRePEc is based on the data model and the basic technology developed at SocioNet project, it is considered to be a CERIF compatible CRIS.

This paper as a first in the planned series has its main focus on mapping a new area for CRIS development. A set of the reasons for developing new scholarly communication tools is mostly analyzed and discussed in here. More technical details about SocioRePEc are expected to be presented in the following papers.

In the next section the traditional mode of a research outputs (publications) usage provided by a system of academic publishers and journals is analyzed. Also the shortcomings of this mostly “one-way” usage mode and a “passive” character of the related scholarly communication are discussed.

The third section highlights some new possible options of the research publication usage, which can be designed in CRIS. The paper concludes that this type of the advanced usage options can initiate pre-publication scholarly communication.

The fourth section provides general details about SocioRePEc CRIS including implementation of the new options of research outputs usage and pre-publication scholarly communication. These new CRIS features provides researchers with the interactive mode of the research outputs usage.

2. Global scholarly communication mechanism

In general, the global scholarly communication mechanism is illustrated by the Figure 1. In the traditional scholarly publishing system researchers find and read different types of research publications (outputs). While reading they select from the text some useful pieces of information. We can call these selections as the “research artifacts”. Researchers mentally manipulate with the artifacts to find out scientific relationships among them. They then create their own research artifacts. To share the created relationships and artifacts with the community, researchers have to use the scholarly publishing system. It means that researchers have to compose and present the newly created relationships and artifacts in a form of a canonical research manuscript, like working paper, article, book, etc. with citations, references and some other mandatory attributes. They submit manuscripts to publishers.

After publication, researchers typically are monitoring the publications' use. They may be waiting some occasional comments on it. And this cycle as shown at Figure 1 is repeated by researchers again and again.

Scholarly communication supported by the traditional scheme, as at Figure 1, has obvious weaknesses and shortcoming.

While manipulating with artifacts researchers make with them some mental experiments to combine them with their own research assets. This is some kind of “trial and fail” actions. As a result of such “trial and fail” experiments researchers either accept some artifacts by connecting them with their own research assets or reject them. In both cases, researchers eventually get some individual experience about the scientific value and/or the potential impact of these tested research artifacts. This information can be very useful for authors of the tested artifacts and for the community at large. But currently only a small part of this information becomes public.

In the traditional scheme of the global scholarly communication most authors hardly get any feedback on their papers. They usually don't know if, let alone when, someone try to use their research results and fails. If someone really use the authors' results and cite them in their own papers, the authors had no chance to assist them on the proper usage of the results, to correct a wrong usage or to get a better effect from using their results. From a point of cooperation inside a scientific division of labor, this situation looks inefficient.

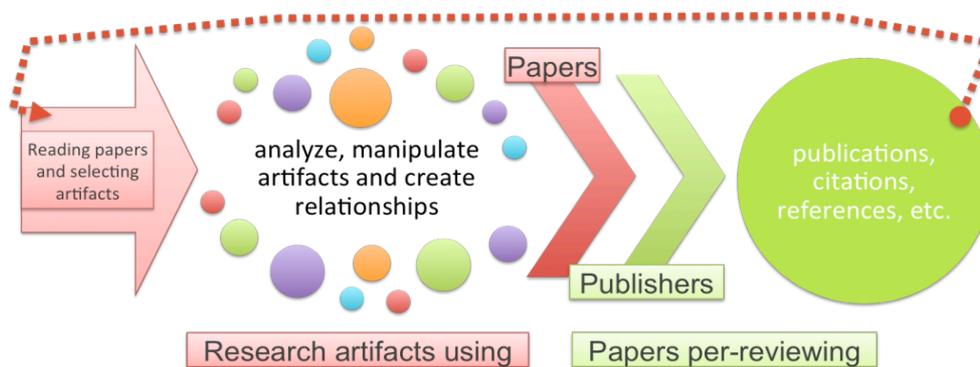


Fig. 1. Traditional scholarly communication scheme through academic publishers and journals infrastructure.

Generally, researchers manipulate not research publications but the research artifacts contained in the publications. They also manipulate the relationships between artifacts. Why they could not share these artifacts and relationships as their micro research outputs with the other researchers? Scientific circulation of such micro research outputs [9] along with canonical papers can give the community multiple benefits, like better transparency in research, a rise of the traditional publications' quality, and so on.

3. Pre-publication scholarly communication

The main challenge of improving transparency in research is a public visualization of the significant part of the process of the researcher's scientific analysis and creativity. This part is traditionally hidden from the research community, because it is carried out in the mind of the researcher. At the Figure 1, it is the area “analyze, manipulate artifacts and create relationships”.

To meet this challenge, the research community needs a CRIS that allows scientists to publicly express what kind of the research results (artifacts) they select as interesting and which ones they use in their work, and later, how these results have been used in their study. The traditional technique of the scientific citations does not help much with this, because it operates with the entire publication, but not with fragments. Within this traditional style it is also not really convenient for a researcher to explicitly indicate the motivation for the usage/citation, especially to do it in computer-readable form.

Our vision of the required CRIS assumes that there exists a specific crucial point of global scholarly communication, where appropriate and fortunate changes can notably improve transparency, reliability and

credibility in research. This crucial point for innovations is pre-publication communication between an author of a research micro output (paper) and researchers, who use the results from this micro research output in their own research and cite it in their own research micro output (paper).

Pre-publication communication is initiated by the researchers who use some of the other researchers' outputs for producing their own research outputs.

While reading a paper's abstract at the metadata web page or the full text in PDF file at CRIS content, a user can select text fragments. So the fragment can be selected from either the abstract or the full text. After selecting some text fragment, a user sees a form on a screen. From this form a user can create different types of what we called research artifacts:

1. A user's comment to the selected fragment (both the abstract and the full text). Such comments will be visible to the next reader. A similar user feature already exists in some research information systems (e.g. at ResearchGate.net). There is some infrastructure to aggregate and exchange them (e.g. at Hypothes.is).

2. A research assertion (the same as at Nanopub.org) that a user concludes from the selected fragment (the full text only). Technically, the use of this feature implies a creation of a nanopublication, which can be integrated into appropriated infrastructure [10].

3. A quotation that a user has decided to select from the paper and share with the community as a research artifact. This feature is usually available for the full text only. A quotation is supplemented by a user's comment explaining why it has been selected and shared. Making such quotations the user gets an opportunity to create scientific relationships between some interesting text fragment and other papers, or paper fragments.

4. A micro-preprint where the user's comment is the primary content and the selected fragment is just a base for it. This feature is usually available for the full text only.

All these research usage actions generate private or public information objects (artifacts) that are being stored in CRIS. These artifacts are automatically linked with the initial paper and with the personal profile of the artifact's creator. While reading a paper at CRIS content, a researcher can see all created artifacts linked with this particular paper.

5. There is also a fifth research usage option. It is the creation of a scientific relationship between any pair of research artifacts from the RIS content [11]. It allows to cite any piece of a research output. If a researcher discovers interdependences or relations between any two pieces of research content, (s)he can express these by making scientific relationships. Kogalovsky and Parinov [12] present an initial taxonomy of scientific relationships.

Therefore, pre-publication communication is being initiated immediately after any of these 5 research usage actions have been made. The author of the used/cited research output receives e-mail notification with the data about what the usage/citation means. (S)he can react to these usage/citation actions and the cycle repeats.

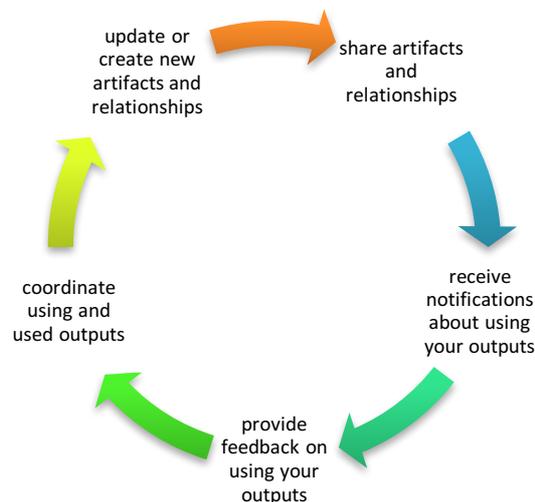


Fig. 2. Pre-publication communication scheme.

The full cycle of pre-publication communication on behalf of an author of some research output, which is used by someone, is illustrated at Figure 2. It works as follows [6]:

1. The author of publication or micro output receives signals (email notification) from CRIS when someone (a user) used his/her research results. Thus communication between the author and the user can be initiated.;
2. Receiving this signal, the author can assist the user in the proper use of the research results. Thus there may be an act of cooperation between the author and the user.;
3. The author can modify the results to enhance the effect of their use. This is also an act of cooperation.;
4. The communication between the author and the user may be published in the scientific information space. In this case the public nature of the exchange may generate competition between different authors to serve the users.;
5. All data are stored in the CRIS and form a public portrait of the researcher. The portrait includes all actions taken and the reactions that allow to evaluate scientific reputation of the researcher. Correspondingly, this significantly increases the responsibility of the researcher for her/his actions in comparison to the situation created by the traditional mechanism of communication.

Pre-publication communication in such a way allows researchers to put into scientific circulation and re-use some fragments, artifacts or smaller pieces of their work, in contrast to a canonical paper.

Generating such artifacts and linking them to each other a researcher can already experiment with the idea that any paper is an aggregation of objects. Thus we can implement the vision of Neylon [9]: “If we take this view of objects and aggregates that cite each other, and we provide details of what the citations mean (this was used in that, this process created that output, this paper is cited as an input to that one) then we are building the semantic web as a by-product of what we want to do anyway”.

4. An interactive mode of the research outputs usage at SocioRePEc

Based on the previously tested approaches [6] and newly created technology [5], a research information system SocioRePEc is being designed. It has recently become available free of charge for the international users at <https://sociorepec.org/>. It is an open non-commercial community project funded by the grants and donations. SocioRePEc is a RePEc-based public research information system. It aims to serve the wide international community of researchers in economics. Many of them are maintaining their personal profiles and research outputs in RePEc. There are more than 50 000 of such RePEc users at the moment. Compared to other RePEc based information systems, the SocioRePEc’s main added value is in the ability to make experiments with the advanced options of research outputs usage and some initial pre-publication communications. The SocioRePEc team is attracting RePEc users and research organization for further development of these tools.

Pre-publication communication, which currently is being developed at the SocioRePEc system, is implementing the interactive mode of their research outputs usage for the system’s users. This interactive mode is a cycle of the following researchers (users) actions:

- a) users can save and store their research artefacts at the system’s data storage;
- b) users can share research artefacts with the community and they share their knowledge or hypotheses about scientific relationships between artefacts and/or papers by linking them using a taxonomy of the applicable relationship classes;
- c) the system allows authors of the linked artefacts/papers to be immediately notified why, where and by whom their outputs were linked and how their research results were used;
- d) the system allows authors of linked artefacts/papers to publicly express their attitudes towards these actions. They can state what they think about the usage of their results;
- e) the system initiates some coordination between a researcher who created an artefact and another researcher who is using this artefact. They can interactively adjust and adapt their “supply” and “demand” to get better mutual effect from their cooperation.

It is important to mention that the basic “author<–>user” communication provided by SocioRePEc is public. That means the system allows experiments with creating additional competition. Researchers who are initially out of the specific “author<–>user” communication can trace the communication and can compete with the “author” by offering the “user” better research artefact or more efficient solution for her/his “demand”.

SocioRePEc allows researchers to make experiments with creating and sharing micro outputs, including relationships. One can imagine SocioRePEc users creating relationships between own outputs and papers/artifacts of other authors, as “neighbors” of the scientific labor division system. SocioRePEc will help researchers to find out who their “neighbors” are. It will initiate scholarly communication between the “neighbors” to get better collective intelligence and to have benefits from direct research cooperation.

As the authors of the Publons fairly wrote: “... in order to build innovative systems for collaboration we need to modify this incentive structure. A corollary of this is that any new approach must augment the current system, at least initially.” [8]. In our case the central idea is to embed the SocioRePEc system with the feature of the interactive usage mode into the traditional communication scheme. The resulting scholarly communication scheme is presented at Figure 3.

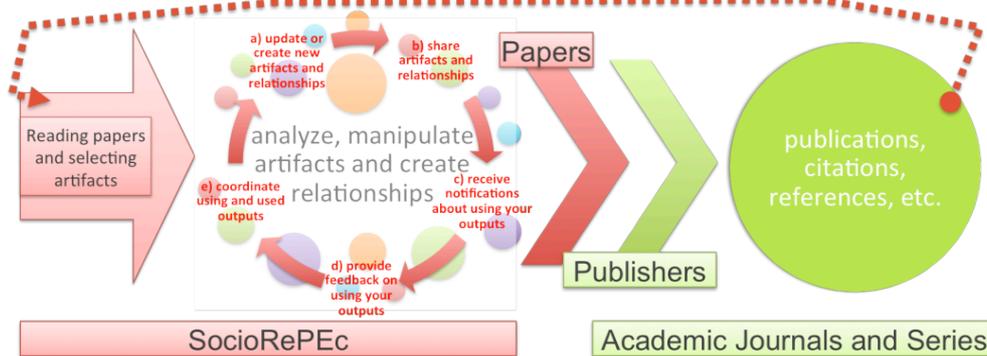


Fig. 3. Scholarly communication scheme as a combination of SocioRePEc facilities and academic journals/publishers infrastructure.

A combination of SocioRePEc facilities with the existing academic journals/publishers infrastructure creates some new opportunities for the research community.

Let’s imagine a hypothetic situation: now scholarly communication is initiated by sharing not publications, but research artifacts and relationships. Hence, researchers don’t need any more formatting their research outputs as papers and making it published in order: (a) to distribute their research results globally; (b) to get feedback on it; (c) to have it used by other researchers in producing new research results; (d) to cooperate with researchers who used their results for getting more mutual benefits, etc.

As a result, it could provide the research community with more efficient mechanism of the global scholarly communication as compared to the traditional academic publishing system.

Acknowledgements

Research results in this paper related with the global scholarly communication technology development are funded by the Russian Foundation for Basic Research, grant 15-07-01294-a. This paper also presents the experimental interactive tools for assessment of non-material assets and conducting applied micro-economic studies funded by the Russian Science Foundation, grant 14-18-01999.

References

1. Parinov S. A CRIS driven by research community: benefits and perspectives. In the proceedings of the *10th International Conference on Current Research Information Systems (CRIS-2010)*, 2010. http://eurocris.org/Uploads/Web%20pages/cris2010_papers/Papers/cris2010_Parinov.pdf
2. Kramer B, Bosman J. Survey of scholarly communication tool usage, 2016. <https://101innovations.wordpress.com/>
3. Parinov S. Open Repository of Semantic Linkages. In the proceedings of the *11th International Conference on Current Research Information Systems (CRIS-2012)*, 2012.

4. Parinov S. Towards an Open Data on how the Research Data are Used: CRIS-CERIF based Approach. In the proceedings of the *12th International Conference on Current Research Information Systems (CRIS 2014)*. 2014
5. Parinov S, Lyapunov V, Puzyrev R, Kogalovsky M. Semantically Enrichable Research Information System SocioNet. In the *Knowledge Engineering and Semantic Web*. Editors: Klinov, P., Mouromtsev, D. (Eds.). Springer International Publishing, 2015; **518**, p. 147-157
6. Kogalovsky M, Parinov S. Scholarly communication in a semantically enrichable research information system with embedded taxonomy of scientific relationships. In the *Knowledge Engineering and Semantic Web*. Editors: Klinov, P., Mouromtsev, D. (Eds.). Springer International Publishing, 2015; **518**, p. 87-101
7. Neylon C. What would scholarly communications look like if we invented it today? *Science in the Open Blog*, 2 SEPTEMBER 2010, Available online: <http://cameronneylon.net/blog/what-would-scholarly-communications-look-like-if-we-invented-it-today/>
8. Preston A, Johnston D. The Future of Academic Research. figshare. Available online: <https://dx.doi.org/10.6084/m9.figshare.871466.v1> 2013.
9. Neylon C. The future of research communication is aggregation, *Science in the Open Blog*, Published: 10 APRIL 2010. Available online: <http://cameronneylon.net/blog/the-future-of-research-communication-is-aggregation/> (accessed on 27 March 2016)
10. Kuhn T, Chichester C, Krauthammer M, N 'uria Queralt-Rosinach, Verborgh R, Giannakopoulos G, Axel-Cyrille Ngonga Ngomo, Viglianti R, Dumontier M. Decentralized Provenance-Aware Publishing with Nanopublications. *Peerj*, 2016. Available online: <https://peerj.com/preprints/1760.pdf> (accessed on 27 March 2016)
11. Parinov, S. Semantic enrichment of research outputs metadata: new CRIS facilities for authors. In *Metadata and Semantics Research Communications in Computer and Information Science 2014*; **478**, p. 206-217, Available online: http://link.springer.com/chapter/10.1007/978-3-319-13674-5_20 (accessed on 27 March 2016)
12. Kogalovsky, M., Parinov, S. The taxonomy of semantic linkages of information objects in research digital library content. *Automatic Documentation and Mathematical Linguistics*, September 2015; **49**:5, p. 163-171, Available online: <http://link.springer.com/article/10.3103/S0005105515050027> (accessed on 27 March 2016)