Global Scholarly Collaboration: from Traditional Citation Practice to Direct Communication
Sergey Parinov, Victoria Antonova

To cite this version:
Sergey Parinov, Victoria Antonova. Global Scholarly Collaboration: from Traditional Citation Practice to Direct Communication. ELPUB 2018, Jun 2018, Toronto, Canada. 10.4000/proceedings.elpub.2018.24 . hal-01816700

HAL Id: hal-01816700
https://hal.science/hal-01816700
Submitted on 15 Jun 2018

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
Global Scholarly Collaboration: from Traditional Citation Practice to Direct Communication

Sergey Parinov and Victoria Antonova

Introduction

1 Current achievements of scholarly digital publishing and citation content analysis create new ways for the better support of the Knowledge Commons. In this paper we discuss the integration of some tools from two ongoing projects. One offers a method of transforming citation data into “interactive elements”, another promises to transform the interactive elements into a communication instrument between citing and cited authors.

2 Both projects are based on the principles of the Open Scholarly Infrastructure (Neylon et al. 2015). This guarantees that even after completing the projects, their outputs will be available as a part of a sustainable infrastructure. It means that if there is a demand for the projects’ outputs, they can be freely re-used and developed by anyone.

3 To discuss how current scholarly communication, based on publishing and citing papers, can be replaced by a collaboration between researchers based on direct communication, we should start with a clarification of the relationships between scholarly communication, cooperation, and collaboration.

4 The main tool of scholarly communication for researchers with the research community is that of publishing their research. References in such publications serve as observable evidence of scholarly cooperation in the research community. Scholarly cooperation, as a kind of socio-economic cooperation (Smith et al., 1995, Axelrod and Hamilton 1981) means that some scholars use the outputs of other scholars and by this, they realize the collective development of scientific knowledge.

5 Citation network analysis allows us to determine whose results the scholar used, i.e. with whom he had a cooperation, and the position of the scholar in this network. For the
The research community as a whole, there is no other common method to indicate what and how outputs were used by particular scholars in their research.

The forms of cooperation used to organize collective activity may vary depending on the number of participants in this activity and other conditions (Smith et al., 1995). Processes of scholarly cooperation based on publishing and citing papers have the form of “horizontal cooperation” (Smith et al., 1995) among the research community.

Scholarly cooperation on a community-wide scale is observable and measurable because:

a. current rules and regulations, which oblige scholars to specify their use of previously published research outputs in a prescribed manner;

b. tools and services for research papers preparation and publishing;

c. means for processing the full text of publications and visualizing citation data, relationships and statistics.

As a result, the research community has well-developed technology to produce and share publications with citations as an instrument of cooperation. There are also formal/informal rules and social institutions that regulate the application of this cooperation instrument.

Let us compare: scholarly cooperation based on the exchange of publications and citations with another well-known form of collective activity based on direct communication between members of a small group like a research laboratory or a project team.

In the first case, the “cooperation is accomplished by the division of labor among participants as an activity where each person is responsible for solving a portion of the problem” (Power 2017). If we look at how a small group works, as in the second case, we see a collaboration, which “is a coordinated, synchronous activity that is the result of a continued attempt to construct and maintain a shared conception of a problem” (Power 2017).

Such comparison shows that publications as a communication and cooperation tool offer participants significantly less effective cooperation. In particular, the authors of research publications do not know about all the facts when someone uses their published research. As a rule, they do not have accurate and complete information on who and when used (cited) their publications, and what exactly was used (cited), and for what purposes. There are no common ways for authors to publicly respond to how other researchers used (cited) their publications. The research community does not have a complete picture of how the publication was used to create new scientific knowledge. Similarly for the authors: it is not known exactly which of their outputs were used, by whom and how exactly.

If we consider citations as a reflection of scholarly cooperative links, then in real academic life the citations bear a lot of information, not all of the same weight. These aspects, including scholars’ lines of behavior associated with citations, and the actual content of citations are studied by the sociology of citations (Adler et al., 2009). As the main type, the “grateful” citations (Adler et al., 2009) are pointed out, considered as an acknowledgment of the intellectual debt in relation to the cited publication. Such citations directly indicate the scholarly cooperation between citing and cited authors.

Some researchers, however, consider (Cozzens 1989) that for modern research publications the “rhetorical” citations (Adler et al., 2009) are more common. They work as
the means of conducting scientific discussion, serve as illustrations, and also perform certain “ritual” functions, not directly related to the process of collective creation of scientific knowledge and scholarly cooperation.

14 This situation is further aggravated by the use of citation indexes to assess research performance. Measuring the success of scholars by the number of citations of their publications affects the essence of the citation process, because in many cases this indicator becomes the goal of scientific activity. Consequently, all the characteristics of the citation process cease to be trustworthy (Neylon 2017).

15 Despite this, the publication of research results and the practice of citation are mandatory in the research community. This allows us to suggest that the data on citation networks extracted from research publications carry information about a substantial part of the existing global network of scholarly cooperation. The processing of these data makes it possible to visualize this network and presumably identify collaborative scholars.

16 Identifying pairs of scholars, citing and cited, one can suppose that they are related by scholarly cooperation. If the contact information of both is available, it is possible to organize direct scholarly communication between them. When such communications becomes possible between all pairs of cooperative scholars, it promises them better research performance.

17 The development of a technology for scholarly cooperation between citing and cited authors creates the conditions when researchers can identify the “suppliers” of the research results required by them, and the “consumers” of their own results. Hence, scholars are able to simultaneously coordinate their research in two directions with the “suppliers” of the research results required by them, and with the “consumers” of their own research.

18 The next step is the creation of the conditions for the emergence of full-fledged collaboration between citing and cited scholars, which normally arises only in small groups.

19 The second section of the paper considers some research information systems (RIS), the combination of which allows the citations found in research publications to be turned into interactive elements. Such innovation, in its turn, allows RIS to initiate direct communication between cited and citing authors.

20 In the third section, we discuss how the small-group mechanism of collaboration works and what RIS functionality is required in order to create conditions for the operation of such mechanism on the scale of the entire research community.

21 In conclusion, the possible consequences of cooperative mechanism implementation for the scientific community as a whole are briefly discussed.

From citing of papers to direct scholarly communication

22 There is a clear motivation to create a technology which allows direct scholarly communication between researchers who currently can only cite each other’s papers. As expected, such an opportunity for scholarly cooperation can greatly increase the research performance of these scholars.
This technology is a combination of existing and newly created tools. One of them is a parsing of citation data from research papers, which also gives us the citation contents around the in-text citations. Within a RIS, the citation content can be used to initiate direct scholarly communication between citing and cited authors. A citation content analysis also helps to use such communications to develop research cooperation.

Parsing of citation data from research papers

Citations in research papers are an important source of data about cooperation between researchers. To analyze this information one has to first extract citation data from research paper content.

The project CyrCitEc (https://github.com/citeccyr) is developing tool for the parsing of citation data, including a context for in-text citation from research papers full-text in PDF (Parinov 2017). The project is funded by the Russian Presidential Academy of National Economy and Public Administration (RANEPA, http://www.ranepa.ru/eng/). It takes input data from RIS RePEc (http://repec.org/) and Socionet (https://socionet.ru/).

In late March 2018, CyrCitEc processed 203 collections of papers with 89,342 publications in total. The largest part of this set are 157 Russian academic journals covering different academic disciplines and provided by NEICON (https://socionet.ru/collection.xml?h=spz:neicon&l=en). There are also research paper series in Russian and English provided by Russian universities, including the Higher School of Economics (https://socionet.ru/collection.xml?h=repec:hig&l=en), RANEPA (https://socionet.ru/collection.xml?h=repec:rnp&l=en).

The approach used by CyrCitEc for citation data parsing was presented in (Parinov 2017). All citation data extracted by CyrCitEc project are publicly available at http://peren.openlib.org/. Regularly updated CyrCitEc statistics about parsing results are available at http://citru.repec.org/stats.html.

Only 69% of the papers have full text PDFs available for the citation data parsing and only 51% of the papers have a list of references in more or less standard form.

Based on the set of papers with references we parsed in total 801,318 references, that is on average 18 references per paper. In this set, about 5% of references are duplicated, because different papers can cite the same publications and have the same references.

For 26,467 of the parsed references we were able to create citation relationships between citing and cited papers, since we found cited papers’ metadata within RePEc and Socionet information systems.

Using the same set of papers and approach, we parsed 750,607 in-text citations. They mention 1,072,175 parsed references. This is 270,857 references more than total number of parsed references, since some references are mentioned more than once. On average, there are 1.3 mentions per reference.

Non-mentioned references were also counted: 110,340 references (14%) have no in-text citations at all. About 37% of papers with references have at least one non-mentioned reference.
In-text citations

Among different types of citation data available in research papers, in-text citations may keep the data about the character of cooperation between citing and cited authors.

Parsed by CytEcCyr the in-text citation data include the following attributes (see also an example of a data record below):

1. a text string of the style of in-text citation, e.g. a number or an author name in square or round brackets (the tag <Exact> in the example below);
2. a link to a reference, mentioned in this in-text citation (the tag <Reference> below);
3. text coordinates of the in-text citation, i.e. a serial number of the first and the last in-text citation symbols counting from the beginning of the paper’s content (tags <Start> and <End>);
4. citation contexts located on the left and on the right of the in-text citation; these include at least 200 symbols expanded for taking a whole sentence (tags <Prefix> and <Suffix>).

An example of parsed data about one in-text citation:

```
<intextref>

    <Prefix>... countries and Soviet republics</Prefix>

    <Suffix>; Gokhberg, Kuznetsova, 2011] ...</Suffix>

    <Start>8757</Start>

    <End>8781</End>

    <Exact>[Gokhberg et al., 2009]</Exact>

    <Reference>20</Reference>

</intextref>
```

Source: https://goo.gl/Eo1FgG

The in-text citation from the example above has a link with a reference having the number 20 in a paper. CyrCitEc parsed for this reference following data:
All CyrCitEc data with in-text citations and related references are input for the next tool, which is a part of the Socionet RIS.

**Socionet tools**

Socionet services, as described in (Parinov 2017), use the in-text citations and references data to produce computer-generated annotations to the content of PDF papers. Fig. 1 shows what these annotations look like using in-text citation and reference data from the examples above.

**Fig. 1. An in-text citation as an interactive element**

Compared with 2000, the higher education sector's share (Figure 2). Still, in this regard Russia remains far behind not only to the university-based R&D model, but even behind some of the republics (Gokhberg et al., 2009; Gokhberg, Kuznetsova, 2010) with businesses does not look very impressive either (Figure 2). However, the current Russian situation with universities' research activities... (Els detected: 2000, 2009, 2010)

Source: https://goo.gl/bZJwzZ

Readers of PDF papers see the in-text citations, if they exist, as annotated text. At Fig. 1 there are mentions of two references in brackets. These highlighted in-text citations work as interactive elements, since clicking on them opens an information box (at the right side on Fig. 1) with additional data about the cited paper. The additional data can include details about the cited paper (citing statistics, title, authors, etc.) and links to additional tools.
One of the Socionet features is the multiple semantic relationships between information entities (Parinov 2013). It allows the linking of citation data with other types of information. A fragment of the semantic linkage network is presented at Fig. 2.

Using these linkages, we can associate additional data with the citation data, e.g. the contact and affiliation data of the authors for the cited and citing papers, metadata of cited and citing papers, etc.

As we expect, in 2019 CyrCitEc will processes all available publications from RePEc and Socionet. As a result, the parsed citation data will be automatically linked with about 70,000 author profiles (https://socionet.ru/collection.xml?h=repec:per:pers&l=en). These author profiles are linked with papers, and with about 15,000 profiles of organizations (https://socionet.ru/collection.xml?h=repec:edi:inst&l=en) that also have links with other author profiles from their staff.

The system, using the linkages, also can recognizes types of relationships between signed-in users and information objects:

1. a user is the author of the browsed object (paper, citation, etc.);
2. a user is a co-author of the selected author;
3. a user is a citing or cited author;
4. a user is just a reader, he/she has no relation with the browsed object.

This information allows the system to initiate different scenarios of direct communications (Kogalovsky and Parinov 2015) between cited and citing authors.

Fig. 2. Semantic linkage network, a fragment

The system can notify an author about new citations of his/her papers, including its context. Well-known RIS ResearchGate, Academia.edu and others already do it. The system can also provide the author with links to PDFs, which cite his/her papers, where the new citations are highlighted/annotated and work as interactive elements or as a communication instrument.
If the system identified a user as the author of the cited paper and this user clicks the in-text citation pointed to his cited paper, the system allows this user to express publicly or privately his reaction on how the citing author used/cited his research. According to the citation content, this reaction can be simple as “agree/disagree”, or can provide explanations of the cited author what was wrong with using his outputs, or how it could be used properly, etc. If the cited paper has several co-authors the system allows them to express their “agree/disagree” with a reaction of one of them.

The system itself also can initiate some direct communications between citing and cited authors using the sense of the citation content, such as citation polarity or citation function.

Citation content analysis

Citation content analysis, for which CyrCitEc already provides about 750,000 records of the right and left contexts for each in-text citation, has a lot of attention of researchers. Waltman (2016) in his review of the traditional citation impact indicators proposed different ways to improve the indicators, including taking into account “the context in which a publication is referenced (i.e., the sentences in a citing publication around the reference to a cited publication)” (Waltman 2016 p. 43).

In recent years, methods for analyzing the content of citations have been actively developed. Some studies (Zhang et al., 2013; Ding et al., 2014) present concepts of content-based citation analysis (CCA), which addresses a citation’s value.

Practical experiments with the analysis of in-text citations (they are also called as the in-text references) on various sets of full text papers are also known. One of them identified verbs in citation contexts (Bertin and Atanassova, 2014) and later they characterized the different sections of articles in terms of the verbs that appear in citation contexts (Bertin and Atanassova, 2015). Another aspect of CCA is how references are distributed along the structure of articles, and the age of these cited references (Bertin et al., 2016). Some authors analyzed in-text citations as functions of time, textual progression, and scientific field. They built characteristics of in-text citations in over five million full text articles (Boyack et al., 2018).

Hernández-Alvarez and Gómez (2016) in their survey of CCA provided information about tasks, techniques, and resources, including such tasks as the citation polarity and function classifications.

The analysis of citation polarity/function has the potential to draw conclusions about the motives of authors in citing papers. Such analysis can also produce suggestions: what exactly and was used from the cited papers and why. In some cases, this information may be critically important to the authors of the cited papers and may help to initiate direct communication between them and the citing authors.

If CCA recognizes criticism and the system notifies the author of the criticized paper, it gives him an opportunity to correct mistakes and further develop his research.

If CCA informs the author about a positive impact of his/her paper, then the author can conclude how to develop his research to strengthen research results of other scholars who cited them.
In the both cases, cited authors will benefit if they inform citing authors about their progress with the cited research.

As a result, citation networks will become true communication networks. When implemented, such a system should theoretically allow researchers to directly collaborate with each other without the mediation of the current publishing infrastructure.

Such and many other possible types of direct communications between a pair of researchers, where one is the cited and the citing, are the driving force of research communication and cooperation development.

However, it is not just pairs of communicating researchers. It is an act of, at least, the triple direct communications. Because, if we observe a citing author, he can have direct communications with authors who cited him.

We should address the situation when researchers cooperate and coordinate their activity in a group that includes three parties:
1. themselves (“producers”)
2. authors whom they cite (“suppliers”)
3. authors who cite them (“consumers”).

It means cooperation where “suppliers” and “consumers” also have direct communication and can directly affect each others’ activity. Since the group is “a collection of people committed to work jointly toward at least one group goal” (Randrup et al., 2016), the goal of this group is obviously the creation of new scientific knowledge.

Towards a global scholarly collaboration system

By providing direct scholarly communications for the participants of scholarly cooperation, who have traditionally collaborated via publication exchange and citation, it becomes theoretically possible to create more favorable conditions for their collaborative creativity and the development of new scientific knowledge.

Let us consider a joint activity of scholars in small groups (i.e. laboratories, project teams, etc.), which are strongly based on direct communication between group members.

A group, and specifically a small group, is “a distinguishable set of two or more people who interact, dynamically, interdependently, and adaptively toward a common and valued goal/objective/mission, who have each been assigned specific roles or functions to perform” (Mathieu et al., 2000, p. 274). A key feature of small groups is a harmonization of activities on the principle of “all with everyone.”

Another important feature of small groups is the high variability of their environment and, as a consequence, the need for group members to quickly adapt to changing conditions. As Mathieu et al. underlined, “… in order to adapt effectively, team members must predict what their teammates are going to do and what they are going to need in order to do it” (Mathieu et al., 2000, p. 274).

Cooperation is generally defined as a “joint effort toward a group goal” (Randrup et al., 2016) and as a “concerted collaboration”, which is a collaboration with “no identifiable individual deliverables; only group deliverables, toward which members must contribute simultaneous efforts” (Randrup et al., 2016).
Cooperation as a collective activity of people becomes concerted, if the participants can create and constantly update the collective model of their activity and habitat. They need this model to “play” (simulate) and analyze the various possible options for cooperation. Such a collective model arises if the group members can share with the group their personal mental models and hence create a collective (team) mental model.

The basic idea of mental models is that humans by their mental reflection construct internal working models of the world. “When interacting with the environment, with others, and with the artifacts of technology, people develop internal mental models of themselves and the things with which they are interacting. These models provide predictive and explanatory power for understanding these interactions” (Badke-Schaub et al., 2007, p. 7).

The concept of a shared or collective mental model is defined as “knowledge structures held by members of a team that enable them to form accurate explanations and expectations for the task, and, in turn, coordinate their actions and adapt their behavior to demands of the task and other team members” (Jonker et al., 2011).

The basis of the collaboration mechanism, which works for a small group, is that of shared mental models. A background in this area includes research on the development of public institutions (Denzau and North, 1994), increase in the effectiveness of joint activities of people in a group (Mathieu et al., 2000), interaction of people with software agents (Fan and Yen, 2007), environmental protection (Jones et al., 2011), political activities (Richards, 2001), etc.

The sharing of group members’ mental models means that members inform each other about their intentions and possibilities regarding options of their joint activity. An aggregation of such information, received from all members, forms a choice area, which is available for analysis to each individual member.

“Shared or team mental models are characterized as knowledge or belief structures that are shared by members of a team, which enable them to form accurate explanations and expectations about the task, and to coordinate their actions and adapt their behaviors to the demands of the task and other team members” (Badke-Schaub et al., 2007 p. 8).

In (Parinov 1999) we proposed a conceptualization of how a collaboration mechanism based on shared mental models works.

An aggregation of members’ shared mental models creates a collective mental model of the team. Team members interact with the collective mental model by taking information from it, playing with it (sorting out different ways of their cooperation), and by changing it. They can change in the model their personal information image. They can also propose new configurations of the group’s cooperation.

In Fig. 3 we illustrate these interactions by the example of forming a collective mental model (CMM) for a group of 4 members. Each member, by continuously exchanging information with others, forms and actualizes his/her own mental model of group cooperation and alienates it into collective model for decision making about the future group’s activity.
If group members have fixed in CMM a mutually acceptable configuration of their cooperation, then this configuration passes to the stage of practical implementation. The CMM, formed and set at this stage, is used by the members for the actual coordination of their practical activities.

There is a lot of literature on collaboration research, such as the Six Patterns of Collaboration (Briggs et al., 2014), which suggests a conceptualization of collaboration as the following processes: generate, reduce, clarify, organize, evaluate, and build commitment. The same authors also provide the Six-Layer Model of Collaboration (Briggs et al., 2014), which includes collaboration goals, group work products, group activities, group procedures, collaboration tools, and collaborative behaviors. Such research claims an intellectual foundation (Randrup et al., 2016) for discussing computer-supported collaboration, collaboration support systems (Briggs et al. 2013), integrated collaboration environment (Vindasius 2008) and many others.

However, they do not address the basic collaboration instrument, that is, shared mental models. Without it, it is impossible to respond to the obvious research question: how can the mechanism of scholarly collaboration, which traditionally serves only members of small groups, work for the entire research community?

In this paper, we do not claim to give an exhaustive explanation of the question posed. Below we discuss what main tasks should be solved to make CMM a part of the social and technical research infrastructure. One of possible options is to implement CMM in RIS, like RePEc or Socionet. Based on services of these RIS we can implement CMM to develop...
tools for direct communication between citing and cited authors. Since these RIS for economics have about 70,000 identified authors linked with their papers, the proposed mechanism of collaboration will serve a significant part of this research community.

A note: CMM within a RIS already is not a “mental” model. In that case, it would be better to call it a collective information model (CIM).

On a way to having CIM as a part of social and technical research infrastructure we suggest the following main tasks:

1. A synchronization of the individual mental models of the members of the group with their information images in CIM.
   This function can be realized with the help of an information system that collects and accumulates various data about the activities, intentions and capabilities of cooperative researchers. Tools need to be developed that allow a person to share his mental model with the information system, and also to ensure its continuous realization.
   The parameters of the implementation of this function are: the number of group members reflected in the CIM; the accuracy and completeness of the representations of their behavior, intentions and possibilities; and the speed and accuracy of updating changes in these data.

2. Representation in CIM of the environment in which the group works, and a reflection of the changes taking place in this environment.
   An information system can also perform this function by collecting information about the environment where the group members cooperate. It is possible that in the near future this function will be implemented even more effectively in connection with the development of the “internet of things”.
   The parameters of the implementation of this function are: the size of the fragment of the environment reflected in the CIM; the accuracy and completeness of the representation of the corresponding fragment of the environment; and the level of actualization of changes in the environment.

3. Playing (making simulations) in the CIM of possible variants of group members’ cooperation.
   Various computer simulation models can be useful to implement this function, allowing a computer analysis of the best scenarios.

4. Choosing the best variant for cooperation from many possible.
   With equal relations between group members (no subordination), the implementation of this function means that the members must negotiate. This requires a mechanism of a collective decision-making.

5. Realization of the chosen variant of cooperation in practice, including management functions over the joint activities of the group members.
   This requires a mechanism for collective management of joint activities.

Summing up listed above, we can conclude that the implementation of the first three tasks are greatly influenced by a power of information technology, while the last two tasks also depend on social norms and rules, i.e. on institutions of cooperation.

**Conclusion**

If the small-group mechanism of collaboration is used on a larger scale, a cooperation in the research community may works in more effective mode. Cooperative scholars can faster coordinate their activities by direct communication that should give them better research performance.
Another consequence is that traditional publications and academic publishing infrastructure lose their monopoly as an instrument of global scholarly communication. This will create some challenges for sustainability of global research community.

Another serious challenge is the rapid increase of communication among cooperative researchers resulting in a danger of information overloading for them. As Randrup et al. wrote: “Core insight with a significant negative impact on the performance of collaboration is which have been unveiled by research is cognitive overload and inertia. Individuals have limited attention resources” (Randrup et al., 2016). It is a threat, but current research and development in areas like software agents, computer bots and artificial intelligence help humans to cope with an increasing intensity of information flows and give optimistic perspectives for surviving in a coming digital era.

BIBLIOGRAPHY

References


Neylon, C. (2017). “Citation metrics are making headlines, but what does citation really mean?” *JISC blog*: https://jisc.ac.uk/blog/citation-metrics-are-making-headlines-but-what-does-citation-really-mean-08-dec-2017


http://webuser.bus.umich.edu/sja/pdf/Intra-InterOrg.pdf


ABSTRACTS

The development of recent research information systems allows a transformation of citations in the full text of research papers into interactive elements. Such interactivity in some cases works as an instrument of direct scholarly communications between citing and cited authors. We discuss this challenge for research e-infrastructure development including opportunities for improvements in research cooperation and in collaboration mechanisms for the global research community.

INDEX

Keywords: Citations as interactive elements, scholarly communication, cooperation, collaboration, shared mental models

AUTHORS

SERGEY PARINOV
CEMI RAS, RANEPA, Russia
sparinov@gmail.com
(corresponding author)

VICTORIA ANTONOVA
Higher School of Economics, Russia
v_antonova@mail.ru