A Challenge of Research Outputs in GL Circuit: From Open Access to Open Use

Draft, November 14, 2013

Sergey Parinov¹, Mikhail Kogalovsky² and Victor Lyapunov³

1) Central Economics and Mathematics Institute of Russian Academy of Sciences (Nakhimovsky pr. 47, 117418, Moscow, Russia)
2) Market Economy Institute of Russian Academy of Sciences (Nakhimovsky pr. 47, 117418, Moscow, Russia)
3) Institute of Economics and Industrial Engineering, Siberian Branch of Russian Academy of Sciences (pr. Lavrentieva 17, 630090, Novosibirsk, Russia)

Acknowledgments

We are grateful to Keith Jeffery and Cameron Nylon for useful comments on the “open use” concept. This research is partly funded by Russian Foundation for Basic Research, grant 12-07-00518-A.

Abstract

Open Access movement and currently formed GL circuit provide the scientific community with unique opportunity to modernize a fundamental part of research life-cycle: processes by which the scientists reuse research outputs when they produce new knowledge and then the community assesses their impact. When scientists mentally manipulate the research outputs, outcomes and other objects of scientific information space they discover relationships between the objects and thereby they reuse it to produce a new scientific knowledge. Some of these relationships become visible in scientists’ articles (e.g. by citations). Most of them are directly not observable and may exist in a mental form only. In the paper we propose an “open use” approach for the research area and discuss a practical implementation of the model within a research information system “Socionet” supported a grey literature circuit.

Introduction

Cameron Neylon wrote (Neylon 2012): “open access must enable open use” that means the “innovators can manipulate the material” and from the technical side it requests a “standardizing the representations of data and knowledge in ways that make them easily transferable”.

From 2000 we are building a step by step for Russian language scientific community a research information system Socionet\(^1\) that implements a combination of the open access and open use approaches.

From the beginning Socionet supports the grey literature circuit. It provides tools for electronic depositing and distributing different types of research materials in a way not controlled by commercial publishers. Socionet services make a standardizing representation of deposited materials within the research data and information space in a way that makes them easily transferable. Socionet users have a personal information robot service, which notifies them about new materials relevant to their interests. At Socionet there is a statistical subsystem which collects various data about scientists’ activities in this virtual research environment and produces publicly available and daily updated scientometric indicators.

In the first section of the paper we present our “open use” approach and in the second one – some technical details about the Socionet research information system and about the current state of this approach implementation within this system.

**An “open use” approach for research area**

The proposed “open use” approach is based on a following assumption: researchers use available research outputs to produce new scientific knowledge when they mentally manipulate the research artifacts, extracted from the reading materials, and discover scientific relationships between the artifacts and their own outputs. Some of these relationships become visible as citations in researchers’ outputs.

\(^1\) http://socionet.ru/
Another part of the relationships is directly not observable since the existed citation technique does not allow researchers to express them explicitly and correctly.

Information about some relationships remains in a mental form only. As a result it is not shared with the research community, it is not utilized in a global research process, and the community has no complete picture about scale and scope of research outputs using and impact.

To respond on this challenge we are developing a concept and building a digital technology of "open use" versus traditional paper-based technology which limitations are mentioned above.

We understand the "open use" for the research area\(^2\) as a process specified by the assumption above, but with at least 5 additions.

1. An open access to research, which is the prerequisite of the open use: all research outputs and full corpus of the Science should be publicly available for using by the community to produce new scientific knowledge. Approaches and a technology to support open access are well known and we do not discuss it here.

2. An openness of results of researchers’ manipulation of the materials. It should be clear specified what pieces of the materials were selected by the researcher as artifacts for its further using.

From technical point it needs an approach allowing scientists to share with the community research outputs in more reusable form than traditional journal articles, books, etc., allow to do it. Currently there are technologies supporting micro-publications (Clark et al. 2013), nano-publications (Groth et al. 2012) or research artifacts (Parinov 2010a, 2010b), which designed to be better reusable. Also there

\(^2\) We consider the open use for research outputs only within the research area and do not discuss in this paper other types of using research, e.g. using of research outcomes, etc.
is an open annotation approach\(^3\) which allows making research artifacts right over electronic version of publications in all traditional forms.

3. An openness of researchers’ motivations to use selected artifacts in producing new scientific knowledge. But only part of researchers’ manipulations with the material leads to real using of research artifacts. So it is important to share with the community also details about not using of artifacts, when there were tries and fails. In that case the result of researchers’ tries and fails in using the artifacts also have to be publicly available.

Technically it can be resolved by implementing semantic linkage technique (Parinov 2012a, 2012b), which in combination with available ontologies allow scientists to express explicitly their knowledge, opinions and hypotheses about scientific relationships between research artifacts and so can visualize in computer-readable form facts and motivations of using or not using.

4. A guaranteed awareness of researchers on all facts of using their research outputs (tries/fails data and motivations) and about impacts of the outputs.

It can be achieved by creating electronic notification system which will trace facts of research objects using and will provide information about this for all interested parties.

5. An openness of usage statistics aggregated by a research output, a researcher (e.g. for all research outputs by this author) and an organization (e.g. for all research outputs produced by staff), including outgoing usage (e.g. how the object used research outputs) and ingoing as well (e.g. how the object was used by the community).

\(^3\) http://www.w3.org/community/openannotation/
Technically it needs a monitoring service, which trace all changes in research objects and semantic linkages among them, collect and process this data to provide public scientometric indicators.

In the next section we present a current (November 2013) state of implementing this approach and building the “open use” technology within Socionet research information system.

**Socionet overview**

The Socionet system development was started in 1997 as a Russian Virtual Laboratory for Economists and Sociologists project. At the beginning it provided a mirror of RePEc.org data and functionality. It also included the first in Russia scientific open archive to submit scientific grey literature in Social Sciences for its online presenting, and some simple tools of virtual workspace (Krichel and Parinov 2002). In 2000 the designed information system got its current name "Socionet" (socionet.ru), since from that time it has own harvester, which federates more research collections and archives, than RePEc provided (Parinov et al. 2003). It allowed a building and, from that time, an everyday updating the Russian research data and information space (DIS) for Social Sciences.

In 2002 a Socionet Personal Zone service was created as add-in online workbench and a managing system for academic electronic assets including the grey literature (samizdat) materials. It allowed a depositing and managing of electronic scientific collections for 9 data types (e.g. "person", "institution", "paper", "article", "book", etc.). The Personal Zone service also included software of the "personal information robot" to trace new additions/changes within DIS according personal research
interests of users and notify them about relevant findings (Parinov and Krichel 2004).

In 2004 Socionet users got some new tools to create and manage semantic linkages between information objects of DIS. From that time some information objects in Socionet, like personal and organizational profiles can represent professional social networks of appropriate research actors (Parinov and Krichel, 2004).

In 2007 monitoring of DIS changes and statistics automated services were started. The Socionet scientometric database has been accumulating from 2007.01.01. The Socionet statistics section provides a big set of time series indicators. It includes indicators of views/downloads aggregated according linkages between DIS information objects, e.g. a sum of views/downloads for all publications linked with a personal profile, or the next step of aggregation – a sum of personal indicators for all people linked with an organization's profile, and so on (Kogalovsky and Parinov 2008). The monitoring service of the Socionet can trace changes in linkages including its semantic. Appropriate scientometrics data is adding to the Socionet statistics subsystem (Kogalovsky and Parinov 2009).

In November 2013 the Socionet system federates more than 4000 collections with scientific materials organized on the base of RePEc.org and about 600 collections from Russian research organizations. In total it is about 2 M materials and with every day average surplus of 300 new materials and 1-2 new collections per week. It covers 15 scientific disciplines organized by 16 data types sections.

From 2009 the Socionet works as a multidisciplinary RIS freely available for all types of academic actors and based on Open Science ideas (Parinov 2009, 2010b). Socionet tends to be a full-functional modern CRIS driven by the community of scientists speaking and communicating in Russian language (Parinov 2010a).
Socionet currently includes following main subsystems (see Figure 1):

1. Information hub (IH), which federates scientific metadata of RIS, RePEc archives and so on. The IH can harvest local metadata organized in different formats. At IH's output one gets accumulated and daily updated metadata in standardized form. Technically IH's output is designed to fit with software agents and give back metadata through OAI-PMH and other XML-based protocols (Parinov 2006).

2. Interdisciplinary research data and information space (DIS) as a visualization of full IH contents presents existed information objects and semantic linkages between them for navigation and searching by Socionet users.

3. Online workbench to create, manage and submit to DIS single materials, whole collections and archives, and also to create/manage networks of semantic linkages between DIS objects. Any authorized researcher or research organization can use it to provide to DIS a proper professional presentation. Profile of organization with linked collections can be represented as OAI-PMH archive⁴.

4. Monitoring and scientometric services, which provides for research community useful scientometric database (updated daily) and notifications. All counted scientometrics indicators are public and can be used for research assessments and scientometric studies.

In November 2013 about 5M semantic linkages exist over research objects of Socionet DIS. At the moment only smaller part of them was created by scientists using Socionet services. About 700 thousands of semantic linkages the Socionet received with RePEc collections (linkages associated with personal and organization profiles). About 4M semantic linkages with the meaning “citation” were imported from the CitEc data base (Barrueco and Krichel 2005).

⁴ see 21 Socionet based OAI-PMH archives at http://roar.eprints.org/view/geoname/geoname=5F2=5FRU.html
Automatically accumulated in Socionet semantic linkages data are used: a) to build a visualization of DIS structure in a form of a graph and to provide graphical navigation tool; b) to search linkages according specified parameters (e.g. by creation/revision date, or by usage characteristics, etc.); c) to create reports for notification system; and etc.

The Socionet system uses the CERIF model of standardizing for the representations of data and knowledge. CERIF Semantics and SPAR ontologies for using within Socionet were converted to a form taxonomy represented by a set of semantic vocabularies. The CERIF based semantic linkage technique after some upgrades allows scientists to link different pairs of information objects from RIS content. The semantic meanings assigned by the scientists to the created linkages carry information about classes of relationships between research information objects.
The scientific relationship classes are defined by taxonomy based on controlled semantic vocabularies produced from available ontologies.

**Implementing “open use” approach at Socionet**

Since approaches and a technology to support open access were implemented at Socionet from the beginning in this section we make focus on new tools and services aimed to support the “open use” approach for the research area.

**New forms of research outputs for better reuse**

Socionet users have an ability to deposit information objects with types “artifact” (Parinov 2010a, 2010b) which has similar functionality with micro-publications (Clark et al. 2013) and nano-publications (Groth 2010). Benefits of depositing research outputs in such new form are (Clark, et. al 2013): “(a) their internal structure is semantically clear and computable; (b) citation networks can be easily constructed across large corpora; (c) statements can be formalized in multiple useful abstraction models; (d) statements in one work may cite statements in another, individually; (e) support, similarity and challenge of assertions can be modelled across corpora; (f) scientific assertions, particularly in review articles, may be transitively closed to supporting evidence and methods.”

**Motivations to use or not use research artefacts**

Socionet users can create semantic linkages between any available research information objects (Parinov 2012a, 2012b). Using available scientific ontologies embedded into the semantic linkage technique users can express explicitly their knowledge, opinions and hypotheses about scientific relationships between research
artifacts and so they can visualize in computer-readable form the facts and motivations of using research artifacts for their research process or not using them.

Available for Socionet users a list of motivations is specified by a set of semantic vocabularies presenting scientific relationship classes which can exist between pairs of research objects of different types.

Initial set of rendered scientific relationship classes has been built from different already existed ontologies (Parinov and Kogalovsky, 2011; Parinov 2012a) includes: (1) relationships between research outputs like inference, usage, impact, comparison, evaluation, etc.; (2) relationships between elements of the set \{scientists, organizations\}; (3) relationships between research outputs on the one hand and elements of the set \{scientists, organizations\} on the other.

Since a semantic linkage expresses a relationship between two objects, there should be determined which scientific relationship classes (semantic vocabularies) applicable for each combination of pairs from a list research objects’ types: a source object type \{"person", "organization", "research output", "project", etc.\} -> a target object type \{"person", "organization", "research output", "project", etc.\}.

The initial classes of scientific relationships and a set of semantic vocabularies were proposed in (Parinov and Kogalovsky 2011). For the pair of object types "research output" -> "research output" following classes of scientific relationships and associated semantic vocabularies were specified (ontologies used as a source for semantic vocabularies are mentioned below in brackets):

- Type "Inference", initial semantic vocabulary (CiTO): "obtain background from", "updates", "used as evidence", "confirms", "qualifies", etc.;
- Type "Research usage", initial semantic vocabulary (CiTO): "contains assertion from", "uses data from", "uses method from", "corrects", "refutes", etc.;
• Type "Hierarchy and association relationships", initial semantic vocabulary (SKOS, SWAN): "broader", "narrower", "related", "alternative to", etc.;

• Type "Research material components", initial semantic vocabulary (DoCo): "duplicate", "revised", etc.

When researchers build a linkage between created own research artifact and some other research artifact and assign to the linkages semantics selected from one of four listed above semantic vocabularies, they express their motivations to use the artifact which the linkage is directed to.

Additionally there is a relationship class “Usage proposal” which is also valid for pair of data types "research output" -> "research output” and has initial semantic vocabulary: “can improve”, “can illustrate”, “can replace”, etc. Using it scientists can share with the community their ideas on what research outputs can be used to improve/develop some other research outputs.

For the pair of types "person" -> "research output" there is a class "Professional opinions" with initial semantic vocabulary (SWAN): "responds negatively to", "responds positively to", "responds neutrally to", etc. Using this class of semantic linkages scientists can express their results of “tries and fails” for attempts to use the research artifacts. They also can protest (the value "responds negatively to") against wrong opinions expressed by other scientists with their semantic linkages, etc.

**Notifications about reuse**

To provide a guaranteed awareness of Socionet users about facts of using their research outputs, including tries/fails data and motivations, and about impacts of
the outputs, we are building an electronic notification system\(^5\), which monitors all changes over a set of semantic linkages between Socionet information objects and send e-mail notifications to users who may be interested in this.

Different types of notifications produced by this service support a scientific circulation/communication by distributing signals about semantic linkages creation/revision. This service notifies:

1. the authors of objects linked by created or revised semantic linkage, just to inform them about this event, let them know about specified semantics and give them an ability to react on this event (e.g. to protest against specified semantic);

2. the author who is changing his/her object (e.g. an article), if the object has linked (cited) in other objects (articles), that by this action she/he can violate have established linkages and/or its semantics;

3. the authors of semantic linkages, if there were changes in objects specified as a source and a target of the linkages, so they should reconsider their linkages and, if it necessary, correct it;

4. the users of research DIS while they are viewing some DIS object (e.g. the readers of electronic articles) that certain semantic linkages made for the displaying source object (e.g. citations in reading text) can be violated because of the target object (e.g. cited articles) was changed, and an author of the linkages has not updated suspicious linkages (e.g. citations).

If the first three types of notification in the list above can be made by e-mail only, the last one should work as warning, that displayed on the screen when it necessary.

\(^5\) Currently, in November 2013, it is on testing stage
Thus notification service creates additional “open use” approach benefits since it improves scientific circulation and communication because it immediately informs scientists about using their research outputs and authors of semantic linkages can receive a feedback on their actions from authors of linked research objects. It also improves global research cooperation because researchers can immediately react on how their research outputs were used by the community. A cooperation can have at least two ways: a support of the third party research where their outputs were used or a protest against of wrong using or their research outputs.

Statistics of reuse

The Socionet Statistics subsystem\(^6\) recently was developed to provide needed openness of statistics about research artifacts reuse. An approach for designing research artifacts reuse indicators was presented in (Parinov and Kogalovsky 2013). At the moment (November 2013) this additional functionality is under testing and evaluation. In Socionet still there is no statistically significant amount of semantic data. Examples of reuse indicators below are provided for illustrative purposes only.

The Figure 2 gives an example of overall statistical distribution of scientific relationship classes (described above) where only three of them classify reuse of research artifacts: "Research usage", "Hierarchy and association relationships" and "Research material components". The Socionet semantic vocabularies also cover "Inference" class of relationships which does not presented at Figure 2 but the same classifies a diversity of research reuse.

\(^6\) [http://socionet.ru/stats.xml](http://socionet.ru/stats.xml)
The Figure 2 illustrates a use case of aggregation of all existed in Socionet scientific relationships between a researcher's profile (one of authors of this paper\textsuperscript{7}) and other research information objects. Total relationships divided on two sets: a) expressed by outgoing semantic linkages, i.e. created by the researcher, and b) ingoing, i.e. created by other scientists with intention to express relationships with research objects belonged to the researcher.

This division on two sets particularly for relationships expressed diversity of research reuse illustrates how the researcher used research artifacts (left column at the Fig. 2) and how the community used research artifacts created by the researcher (right column at the Fig. 2).

![Figure 2. An example of scientific relationship classes distribution for a person](image)

A table on the Figure 3 characterizes the same personal profile as on the Fig. 2 and it illustrates: 1) distributions of outgoing/ingoing motivations in using research artifacts (top row), and 2) distributions of outgoing/ingoing researchers’ sentiments about research artifacts and results to use them (bottom row).

---

\textsuperscript{7} See the source at http://socionet.ru/stat-lnk.xml?h=repec:rus:ecoper:parinov_sergey.56054-1&l=en
Data in the Fig. 3 top row is built as a subset of scientific relationships that classify research using only (4 classes mentioned above), and it is presented on the Fig. 3 by subclasses, i.e. by titles of motivations.

The left column on the Fig. 3 (outgoing motivations) presents a structure of the researcher’s motivations in his using research artifacts. The right column (ingoing motivations) presents motivations of the community to use researcher’s artifacts.

Data in the bottom row on the Fig. 3 represents a distribution of motivations specified by the relationship class the "Professional opinions" and the associated semantic vocabulary. In the case of “outgoing sentiments” (left column) it characterizes a structure of the researcher sentiments resulted from his mental manipulation with research artifacts. The right column with the “ingoing sentiments” demonstrate a structure of sentiments of the community about the researcher’s artifacts.

![Figure 3](image)

**Figure 3. An example of motivation distributions of a person**

Socionet statistics subsystems provides similar usage “portraits” not only for authors, but also for research organizations and, of course, for research outputs.

This 5th element of our “open use” approach also gives benefits for the community.
Such “semantic cloud” in combination with data about groups of research objects linked by certain scientific relationships makes possible a multilayer stratification of Socionet DIS. By this way one can build a “usage map” for scientific areas, disciplines, specific objects, groups of authors, etc.

Since the semantic data can be aggregated by using information about linked objects (Parinov and Kogalovsky 2013), the aggregators can characterize different objects with variation in selected relationship classes or subclasses. It can present e.g. accumulated usage information about research outputs of one author, or a distribution of motivations and sentiments expressed by one scientist, or the same aggregators for a research organization, a scientific journal, an academic publisher, and so on.

Handling some specific classes of relationships or motivations we can make studies for selected groups of research outputs, authors, or scientific disciplines: which research outputs is used in some specific way, e.g. as a background for scientific inference of another research result, what results are claimed to be a theoretical generalisation of another, and many other according our taxonomy of relationships.

**Conclusion**

A grey literature circuit supported by a research information system gets a lot of improvements and researchers – users of the system - have essential benefits when the open access and the open use approaches and tools are implemented in the system.

**References**


