Semantically Enrichable Research Information System SocioNet

Sergey Parinov[1(✉)], Victor Lyapunov[2], Roman Puzyrev[3], and Mikhail Kogalovsky[4]

[1] Central Economics and Mathematics Institute of RAS, Moscow, Russia
sparinov@gmail.com
[2] Institute of Computational Mathematics and Mathematical Geophysics SB RAS, Novosibirsk, Russia
vic@socionet.ru
[3] Institute of Economics and Industrial Engineering SB RAS, Novosibirsk, Russia
prl@mail.ru
[4] Market Economy Institute of the RAS, Moscow, Russia
kogalov@gmail.com

Abstract. Our paper presents a semantically enrichable type of research information systems, which differs from the traditional one by allowing users to create semantic linkages between information objects and to enrich by this the initial content. This approach was implemented as a whole ecosystem of tools and services at the SocioNet research information system, which is publicly available for the research community. Making semantic linkages, the SocioNet users create over metadata from its content a semantic layer that visualizes their scientific knowledge or hypothesis about relationships between research outputs. Such facilities, in particular, open new opportunities for authors of research outputs. Authors can essentially enrich metadata of their research outputs after the papers have been published and have become available at the SocioNet content. Authors can provide comments and notes for updating publication abstracts, data about motivations for citing the papers in the reference lists, research association with newer relevant publications, data about their personal roles and contributions into the collective research outputs, etc.

Keywords: Research information system · Semantic enrichment · Semantic linkage · Scientific relationship · Taxonomy · Controlled vocabulary · SocioNet

1 Introduction

We present a new type of a research information system (RIS). It provides for its users facilities to create semantic linkages between its information objects in a decentralised fashion. In this way, researchers can express publically their professional knowledge, hypotheses or opinions about scientific relationships between research outputs. These expressions make for semantic enrichment of the RIS content.

We define the research information system as a system designed according to (or compatible with) recommendations of the Current Research Information System

P. Klinov and D. Mouromtsev (Eds.): KESW 2015, CCIS 518, pp. 147–157, 2015.
DOI: 10.1007/978-3-319-24543-0_11

(CRIS) [2, 3] and based on a CERIF data model [4]. This approach provides a proper and convenient environment for implementing the semantic linkage technique and enrichment facilities for authors. The growing popularity of the CRIS-CERIF approach improves standardization and interoperability among numerous research information systems that are necessary for integrating its contents in a form of common data and information space (DIS). The use of CRIS-CERIF guarantees that enrichment data created at one RIS can be easily moved, visualized and re-used outside the originating information system at other RIS or at an Institutional Repository (IR).

The demonstrated approach to make a research information system semantically enrichable is based on the following background: a) the CERIF Semantics [4]; b) the web annotation and the open annotation specifications[1] from W3C; c) taxonomy of scientific relationships which can exist between research information objects [12, 13, 14]; d) the specification of registered authors with their personal profiles linked with their publications [8].

We implemented our approach as a whole ecosystem of tools and services at the SocioNet research information system. It is publicly available for the research community at https://socionet.ru/.

Our ecosystem opens new opportunities for authors of research outputs. Authors can essentially enrich metadata of their research outputs after the papers have been published and have become available within SocioNet. Authors can provide data about their personal roles and contributions into the collective research outputs, comments and notes for updating publication abstracts, data about motivations for citing the papers in the reference lists, research association with newer relevant publications, etc.

In the next section, we provide a short historic overview of the SocioNet RIS development during the last 15 years. We also discuss its current state. The third section provides details about the recent results in developing semantic enrichment ecosystem within SocioNet RIS. The fourth section concludes the paper.

2 SocioNet Research Information System

The Socionet system development started in 1997 as the RuPEc project. That project built a computer mirror and Russian language interfaces for the RePEc.org data and its basic services. It also added the first in Russia public open archive to submit research papers in Social Sciences for its online presenting at RuPEc in Russian language and at RePEc in English [7]. At that time, RePEc already had become the biggest world metadata aggregator system for research papers in Economics. To assist RuPEc/RePEc users with monitoring the intensive input flow of new research papers we designed in 1999 a concept of researchers' personal information robots including its ecosystem [6]. By the concept such information robot as an active personal researcher's software agent should filter the input flow of new papers according the researcher's interests and interacts with information robots of other researchers to provide its owner an improved information support.

[1] http://www.w3.org/community/openannotation/

Using the RuPEc background, the concept of the personal information robot [6] and a concept of the research information space for researchers in Social Sciences [13] we designed a research information system called SocioNet [9].

In 2000, following an award from the Ford Foundation[2] we started to run the SocioNet research information system at http://socionet.ru/. From that time onward SocioNet has run its own aggregator service. It aggregates more research metadata collections and open repositories than RePEc provided. It allowed us to build and update a research data and information space (DIS) on a daily basis. At that time, it served social scientists only. Most of its users came from the Russian language research community. We also presented the SocioNet system to the Russian research community as an open technical platform to develop specific scientific information resources and online services [17].

In 2002, we started an implementation of a concept of a permanent researcher's representation. It should allow researchers to collect and permanently represent their research outputs and personal records regardless of changing institutional affiliation [16]. As a first step, we created the SocioNet Personal Zone as an add-in online workbench and a managing system for academic electronic assets. It allows a scientist to deposit and manage different types of collections of research outputs and materials (e.g. "personal profile", "institutional profile", "paper", "article", "book", etc.). The Personal Zone service also included the 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 [18].

In 2004, we implemented at SocioNet some features of a social network. The users got new tools to create and manage semantic linkages between information objects of DIS. At the beginning, it was linkages between personal profiles <-authorship-> publications and organizational profiles <-affiliation-> personal profiles. From that time, personal and organizational profiles at SocioNet can represent in some way the professional social networks of research actors [18].

In 2007, we opened a scientometric section at SocioNet. A special service is monitoring everyday changes of DIS information objects (publications, personal and organizational profiles, etc.) and creates some statistical representation of activities that produced these changes. The SocioNet scientometric database has been accumulating from the start of 2007. The SocioNet statistics section provides a large 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 [5]. In 2009, a monitoring service at SocioNet started tracing all changes of semantic linkages. We added to the Socionet statistics subsystem some indicators based on the semantic linkages data [15].

In 2011, we used the results of this research to guide social network features development at SocioNet [10]. We proposed a concept of the Semantic Linkages Open Repository (SLOR), which implemented at SocioNet enables scientists to express in a

[2] Ford Foundation provided 4 grants for the SocioNet project during 2000 – 2007.

computer-readable form their knowledge, hypothesis or opinions about multiple scientific relationships that can exist between research outputs [11]. Technically it means that we built a multilayer network of semantic linkages over information objects from DIS content.

Researchers typically re-use research objects, thus creating relationships between them by citing. We propose a concept of a research e-infrastructure semantic segment. It allows scientists unlimited re-use of research information systems (RIS) content. The semantic linkage technique implemented at RIS SocioNet provides researchers with tools and services for semantic linking of any pair of research objects, for which metadata are available within SocioNet content. This instrument also allows researchers a decentralized development of semantic vocabularies that guarantee a covering by this technique any new types of relationships.

The conceptual development of SocioNet after 2013 and its new supported use cases are presented in detail at the next section of this paper.

Since 2009, the SocioNet works as a multidiscipline RIS freely available for all researchers. Socionet is a full-functional modern RIS driven by a community of researchers communicating in Russian [10]. Currently, we are also creating the pure English-language version of the system, which will be available for the international research community and RIS developers.

At the beginning of May 2015, the SocioNet system aggregates more than 8000 collections with scientific materials (all types: personal/organizational profiles, publications, theses, etc.). It includes about 7000 collections from RePEc.org and about 1000 collections from Russian providers of research content. The collections in total have about 3.27 million information objects of different types (records). These include about 220.000 personal profiles and about 30.000 organizational profiles. 300 new records arrive on an average day and one or two new collections arrive in a typical week. SocioNet covers 15 scientific disciplines described using 16 data types.

The total number of linkages among the SocioNet information objects is about 9.4 million. On May 10th 2015 after 3 months of testing and experiments with the new facilities for authors to enrich metadata of their publications there are only 211 semantic linkages recently created by registered authors. Among other linkages the biggest groups are: about 7,3 million citation linkages imported from the CitEc database [1]; about 1,3 million “publication” linkages from researchers’ personal profiles to their publications and about 60 thousands “person” linkages from organizational profiles to researchers’ profiles; and etc.

The semantic linkages data aggregated at Socionet are used in the system to:

- build a visualization of DIS structure in a form of a graph and to provide graphical navigation tool;
- generate scientometric indicators based on accumulated semantic data;
- create reports for notification system; and etc.

Currently SocioNet has following main subsystems that provide for data processing (see Figure 1):

1. The Information Hub (IH) aggregates scientific metadata from different data providers and organizations and gives it back in a standardized form (https://socionet.ru/docs/infohubs.htm).
2. The Research Portal visualizes the interdisciplinary research data and information space (https://socionet.ru/portal.html) using the full SocioNet IH metadata content. It presents aggregated information objects and semantic linkages between them for users' navigation.

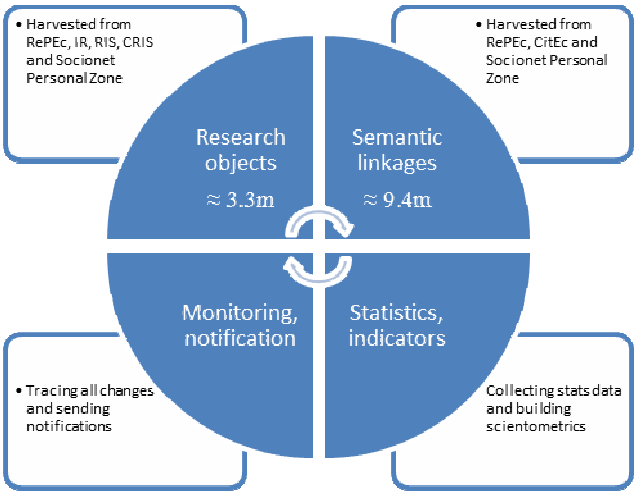

Fig. 1. SocioNet data sources and main services

3. An online workbench called SocioNet Personal Zone (https://spz.socionet.ru/) is there for SocioNet users to create, manage and submit to DIS single materials, whole collections and archives. SocioNet users can create and manage a network of semantic linkages between DIS objects.
4. Monitoring and scientometric services provide researchers useful daily updated scientometric database (https://socionet.ru/stats.xml). All calculated scientometric indexes are public and can be used for research assessments and for scientometric studies.

3 Ecosystem of the SocioNet Tools and Services for Supporting Semantic Enrichment

The first component of the ecosystem is the semantic linkage technique, which allows users to create different types of semantic linkages between papers, personal profiles, etc. This new users' facility is illustrated by the following main use cases:

- Users can annotate text fragments of a paper's abstract to provide readers with additional and/or newer information on the topic. We give details in section 3.1;

- Users can link their papers together, e.g. to provide information about its open access or newest versions; and to show to readers an evolution of ideas or a development of approaches through a set of users' papers. We have more in section 3.2;
- Users can contribute data on how they used the works referenced in their paper. We have more in section 3.3.

Additionally users can specify the roles of their co-authorship in the making of a collective paper, share their professional opinion, or make recommendations, comments about relationships between their papers and the one that is currently browsed by a reader. The system administrator controls all such users submissions.

The second component of the ecosystem is a visualization mechanism. All user-created outgoing and ingoing semantic linkages are visualized on the page that describes a research paper.

The third component is a service for monitoring changes and collecting statistics. All created semantic linkages are processed by the SocioNet system on an everyday base to collect statistical data and to build different scientometric indicators. For personal and organizational profiles, the statistics are aggregated by links between Organization <-> Person <-> Paper. Such indicators are available for every paper, personal or organizational profile.

The fourth component is taxonomy of scientific relationships. We built it current version using available ontologies like SPAR, SWAN, SKOS, etc. [15]. Taxonomy is represented in SocioNet by a set of controlled vocabularies. The research community can create new and develop already existing controlled vocabularies, which are used in tools to create semantic linkages. They can also propose new use cases of using this technique over the research outputs dataset.

The fifth component is the notification service. The registered users can switch on the e-mail notification service, which will inform them when someone creates/modifies a semantic linkage with their papers, or when someone is changing papers that the user linked together, and so on.

The sixth component is the interoperability service. Developers of RIS services can take and use the data of all created semantic linkages. The data are freely available by two main ways: 1) by OAI-PMH protocol with CERIF output format; 2) by REST API in XML form.

3.1 Annotating of Papers' Abstract

After logging into the system, one can annotate the abstract (if it is present) of any paper available at SocioNet. A user can, however, not create a new abstract.

The paper's author makes the annotation of the abstract highlighted by a yellow background colour. If a non-author creates an annotation the annotated text, it will be on a pink background. The text of an annotation is a pop-up. It appears if you point your mouse on the annotated text fragment. All annotations on the paper are listed at the right with links to detailed views of the annotation data. See an example of annotated abstract at Fig. 2.

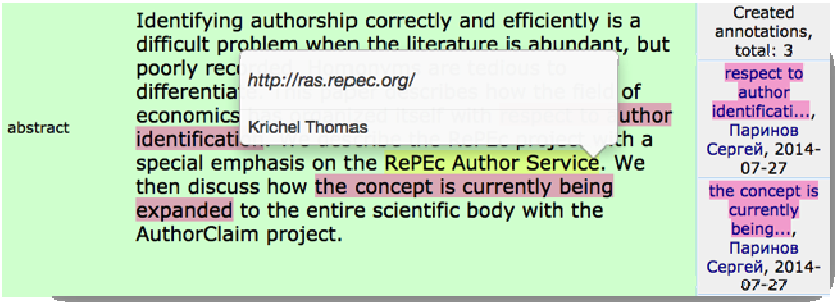

Fig. 2. Example of annotated abstract of a paper

If a user wants to create a new annotation - just select a text fragment within an abstract of a paper. A user will see an icon to open a form to create an annotation.

If a user is an author of the annotated paper, the user will have in the drop-down list "Relationship type" only the blank value "no relationship". It is because the authors cannot evaluate their own paper. But in the other case the user will have in this list a controlled vocabulary called "Professional opinions and evaluation"[3].

From technical point of view all created annotations exist in the SocioNet system as semantic linkages, where the source object of the linkage is the personal profile of the linkage's author, the target object is the annotated paper and the semantic of the linkage is a value from the controlled vocabulary or the "no relationship" value.

3.2 Linking Author's Papers Together

When users are logged into the SocioNet system and browse their papers, they will see to the right of the paper's full-text link the text: "Create a link to a related paper of yours?" Or if the paper's metadata does not have a full text link, there is a text "Create a linkage with appropriate paper?" If the users click on this menu they will get a form for linkage creation as at the Fig. 3.

In this form under the field "Comments", users will see a list of all their claimed publications. A user has to check one of the publications in this list that has a relationship to the paper currently viewed in a browser. Then the users have to select a specific type of the relation between these two papers from the drop-down list the "Relationship type". Optionally they may also enter some comments.

Currently, the system provides two controlled vocabularies for selecting a relationship type between papers belonging to the same author: 1) relationships among versions and components of a research publication; and 2) relationships of development and complement between research outputs.

[3] All used controlled vocabularies are available at https://socionet.ru/section.xml?h=metrics_interdisciplinary&l=en

Fig. 3. An example of a form to create a semantic linkage between papers

Once a linkage is created it is visualized on web pages of both linked papers' metadata as additional information. See an example at the Fig. 4.

Fig. 4. An example of relationships created by an author of linked papers

3.3 An Author Contributes Motives for Using Papers from their Papers' Reference Lists

The SocioNet system uses the CitEc data about citations and similar internal data. When the citation data for a paper is available, SocioNet displays a reference list for the currently browsed paper. The list is limited to publications that are available at SocioNet. In the list there may be two sections: a) referenced papers claimed as own by author(s) of this paper; b) papers of other authors or unclaimed papers.

When users are logged into the system and browse one of their papers, on the right they will see a link "[+]" for each reference. This link opens the form to specify their motives of using the referenced paper.

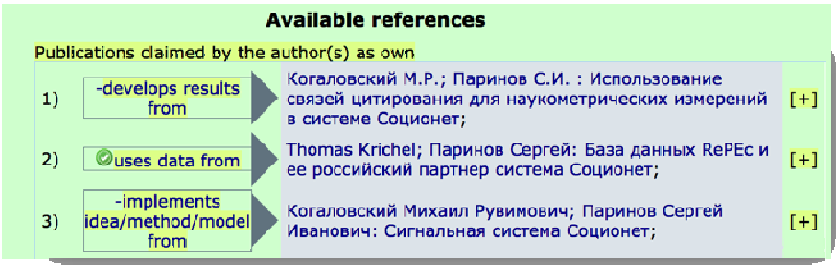

Fig. 5. An example of the enriched reference list provided by the paper's author

For this specific use case, a controlled vocabulary defines relationships of development and complement between two papers, which was used in the previous use case and described in section 3.2.

The data added by the author about citation motivations is highlighted by yellow-colour background as seen at the Fig. 5. Pointing a mouse on it, one may see a pop-up comment if it was provided. The link allows viewing detailed data about the linkage.

4 Conclusion

Since semantic enrichment facilities became available for SocioNet users only in March 2015 we have not sufficient usage statistics at the moment.

The current configuration of the SocioNet system demonstrates potential multiple benefits, which the research information systems of the new semantically enrichable type can provide for the research community.

We believe that the approach we demonstrated has good prospects, because the research outputs authors have at least two reasons to enrich ordinary information about their papers typically provided by publishers. First there is something that is not supported by the traditional publication metadata model (e.g. roles/impacts of co-authors in making a paper, motivations why authors used publications listed at their papers references section, etc.). Second something important may have happened after the paper was published (e.g. authors found some important methodological associations and theoretical hierarchical relations, etc. with research outputs published after their papers).

Acknowledgements. We are grateful to Thomas Krichel for his useful comments. The Russian Foundation for Basic Research currently funds this project, grant 15-07-01294-a.

References

1. Barrueco, J.M., Krichel, T.: Building an autonomous citation index for GL: RePEc, the Economics working papers case. The Grey Journal **1**(2), 91–97 (2005)
2. Jeffery, K., Asserson, A.: CERIF-CRIS for the European e-Infrastructure. Data Science Journal **9** (2010). http://www.codata.org/dsj/special-cris.html

3. Jeffery, K., Asserson, A.: The CERIF Model as the Core of a Research Organisation. Data Science Journal **9** (2010). http://www.codata.org/dsj/special-cris.html
4. Jörg, B., Jeffery, K.G., Dvorak, J., Houssos, N., Asserson, A., van Grootel, G., Gartner, R., Cox, M., Rasmussen, H., Vestdam, T., Strijbosch, L., Clements, A., Brasse, V., Zendulkova, D., Höllrigl, T., Valkovic, L., Engfer, A., Jägerhorn, M., Mahey, M., Brennan, N., Sicilia, M.-A., Ruiz-Rube, I., Baker, D., Evans, K., Price, A., Zielinski, M.: CERIF 1.3 Full Data Model (FDM): Introduction and Specification. euroCRIS (2012). http://www.eurocris.org/Uploads/Web%20pages/CERIF-1.3/Specifications/CERIF1.3_FDM.pdf
5. Kogalovsky, M., Parinov, S.: Metrics of online information spaces. Economics and Mathematical Methods **44**(2) (2008) (in Russian). http://socionet.ru/publication.xml?h=repec:rus:mqijxk:17
6. Krichel, T., Levin, D., Parinov, S.: Active information robot as a software agent of a researcher within RePEc/RuPEc research information systems. In: Proceedings of The first Russian Conference on Digital Libraries (1999) (in Russian). https://socionet.ru/publication.xml?h=RePEc:rus:mqijxk:1
7. Krichel, T., Lyapunov, V., Parinov S.: RePEc and RuPEc: wep-portal of publications in Economics. EL-Pub-99 Conference, Novosibirsk (1999) (in Russian). https://socionet.ru/publication.xml?h=RePEc:rus:ieiets:1999_6863
8. Krichel, T., Zimmermann, C.: Author identification in economics,… and beyond. Working Paper Series des Rates für Sozial-und Wirtschaftsdaten **222** (2013)
9. Parinov, S.: Socionet.ru as a model of the second generation of a research information space. Information Society (1), 43–45 (2001) (in Russian). https://socionet.ru/publication.xml?h=RePEc:rus:mqijxk:7
10. Parinov, S.: CRIS driven by research community: benefits and perspectives. In: proceedings of the 10th International Conference on Current Research Information Systems. Aalborg University, Denmark, June 2–5, 2010, pp. 119–130 (2010). http://socionet.ru/publication.xml?h=repec:rus:mqijxk:23
11. Parinov, S.: Open repository of semantic linkages. In: Proceedings of 11th International Conference on Current Research Information Systems e-Infrastructure for Research and Innovations (CRIS 2012), Prague (2012). http://socionet.ru/publication.xml?h=repec:rus:mqijxk:29
12. Parinov, S.: Towards a Semantic Segment of a Research e-Infrastructure: necessary information objects, tools and services. Journal: Int. J. of Metadata, Semantics and Ontologies **8**(4), 322–331 (2012). doi:10.1504/IJMSO.2013.058415. http://socionet.ru/pub.xml?h=RePEc:rus:mqijxk:30
13. Parinov, S., Bogomolava T.: Creation of the common information space for research communities. In: V International Conference on Digital Publications EL–Pub2000, Novosibirsk, 2000 (2000) (in Russian). https://socionet.ru/publication.xml?h=RePEc:rus:ieiets:2000_6862
14. Parinov, S., Kogalovsky, M.: A technology for semantic structuring of scientific digital library content. In: Proc. of the XIIIth All-Russian Scientific Conference RCDL 2011. Digital Libraries: Advanced Methods and Technologies, Digital Collections, October 19–22, pp. 94–103. Voronezh State University (2011) (in Russian). http://socionet.ru/publication.xml?h=repec:rus:mqijxk:28
15. Parinov, S., Kogalovsky, M.: Semantic Linkages in Research Information Systems as a New Data Source for Scientometric Studies. Scientometrics **98**(2), 927–943 (2013). http://socionet.ru/pub.xml?h=RePEc:rus:mqijxk:31

16. Parinov, S., Lypunov, V., Puzyrev, R.: Socionet System as a platform for developing of research information resources and online services. Digital Libraries **6**, i. 1 (2003). (in Russian). https://socionet.ru/publication.xml?h=RePEc:rus:ieiets:2003_3817
17. Parinov, S., Puzyrev, R., Lyapunov V.: A concept of the permanent online representation of a researcher. In: EL-Pub2003 Conference, Novosibirsk, 2003 (2003) (in Russian). https://socionet.ru/publication.xml?h=RePEc:rus:ieiets:2003_6859
18. Parinov, S., Krichel, T.: RePEc and socionet as partners in a changing digital library environment, 1997 to 2004. In: Proceedings of Russian Conference on Digital Libraries, 29.09 – 01.10 2004, Pushchino, Russia, 2004. https://socionet.ru/publication.xml?h=RePEc:rus:mqijxk:26