Available online at www.sciencedirect.com
ScienceDirect
Procedia Computer Science 33 (2014) 55 – 59
Procedia Computer Science

CRIS 2014

Towards an open data on how the research data are used: CRIS-CERIF based approach

Sergey Parinov*

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

Abstract

Developers of research information systems recently recognized new ambition goals: collecting, processing and making publicly available the data on how scientists use research data to produce a new scientific knowledge. The CRIS approach and CERIF data model can be very useful to respond on this challenge. In the paper we discuss appropriate use cases, necessary functionality and requirements for CRIS-CERIF systems.

Peer-review under responsibility of euroCRIS

Keywords: CRIS; CERIF; research information system; using research data; motivations; use cases and requirements

1. Introduction

A development of research information systems that we are observing in recent years is raising our expectations about positive changes in research practice. Thus Neylon wrote [1]: "… the tools we have to display, manipulate, and interact with this [Open Access] content have become not just incredibly powerful, but easier to use. And as the web in general provides new kinds of services, new ways of communicating, telling stories, and manipulating data there is a profound cultural shift occurring as our expectations of what should be possible, indeed what should be easy, grow".

Some of those expectations are about to have more data on how research data are used by scientists to produce new scientific knowledge [2]. In this specific case the expectations are originated from already existed tools and

* Corresponding author. Tel.: +7-916-914-9051
E-mail address: sparinov@gmail.com

1877-0509
Peer-review under responsibility of euroCRIS
doi:10.1016/j.procs.2014.06.009

services. It includes new approaches to share research outputs: as “nano-publication” [3], “micropublication” [4], “research artifact” [5], [6], etc. Another one is an approach to wrap related research outputs within a “research object” (http://www.researchobject.org/) or “enhanced publication” (http://en.wikipedia.org/wiki/Enhanced_publication). There are also an Open Annotation approach (http://www.openannotation.org/) and the semantic linkage technique [7], which both allow expressing of different types of scientific relationships between information objects, including its non-traditional forms [8].

All these technologies provide an opportunity for scientists to visualize in computer readable form their motivations to use (to cite) available research results when they are creating own research outputs.

Collecting, processing and making publicly available the data on how research data are used are new ambition goals for developers of research information systems. “We need infrastructures that support the sharing of data about the sharing process” [1]. The CRIS approach and CERIF data model can be very useful to respond on this challenge.

In the next section we discuss some possible use cases of these innovative approaches to improve research practice. The third section provides a list of CRIS functionality followed up from use cases and presents our Open (re-)Use concept. In the fourth section we propose a list of requirements to develop CRIS-CERIF models which followed up from discussed topics.

2. New research practice: expectations and use cases

We use an approach developed by the Open Annotation workgroup (http://www.openannotation.org/) as a background, since it provides basic use cases (http://www.openannotation.org/usecases.html) and designs a concept and tools (standardized by W3C, http://www.w3.org/community/openannotation/) for sharing information about using research data as specific case of the open annotating activity. They define the annotating activity is a “creating association between distinct pieces of information” or “annotation is a pervasive activity when reading or otherwise engaging with publications” (http://www.openannotation.org/).

Already existed diversity of research information technologies (including mentioned in the previous section) allows us to formulate some expectations about the future research practice. We present it as a list of use cases (each use case is marked as UC#). Our focus here is only on research practice related with a reusing of research artifacts. The artifact here is some part (fragment, segment, etc.) or a whole research publication.

UC1. A scientist wishes to annotate a scientific relationship that on their opinion exists (as a fact or hypothesis) between a pair of research artifacts.

UC2. A scientist wishes to annotate that he/she used some research artifact to produce a new scientific knowledge. In the annotation he/she wishes to specify for each artifact a motivation to use it, how it was used, how it is impacted on their research. We recognize this case as a demand on the advanced citation technique, which could replace a traditional one for using it within digital research publications.

UC3. A scientist wishes to annotate details on their tries and fails to use some artifacts in their research by linking the annotation with their outputs (as a feedback about loss of citations). It is also can be a part of advanced citation technique.

UC4. A scientist wishes to receive immediate notifications: (1) if other scientists annotated their research outputs (e.g. because they used it in own research); (2) if other scientists changed materials, which were annotated by the scientist (since there is a need to check and a possibility to revise the annotations).

UC5. A scientist wishes to have an ability to react on cases of annotating (using) their research outputs, e.g. to help/assist with proper using of the outputs, or to protest against wrong using, etc.

UC6. A scientist wishes to have a statistical “portrait” (e.g. on some date) of how a research artifact, or a publication, or a set of artifacts/publications of one author/group/organization was annotated (used) by the scientific community. Or how a scientist/group/organization annotated (used in their work) research outputs of other scientists.

Listed use cases are not yet a part of everyday research practice, but it obviously ensures from a nature of research activity [9]. The research community will benefit, if research information systems and particularly CRIS architecture will support such extensions of traditional research practice.

3. New CRIS functionality and the Open (re-)Use concept

The use cases UC1-UC6 from the previous section demand some new functionality of a research information system:

Fun1. Ability to annotate available research materials by using open annotation model (http://www.openannotation.org/usecases.html#requirements) or semantic linkage technique [8] with embedded taxonomy of motivation to use research results, or relationships like scientific inference, hierarchy/association, etc.

Fun2. Ability to target a text fragment within a research material presented a research artifact, which used by a scientist and this fact should be annotated. Ability to deposit such text fragment in a form of micropublication, nano-publication, research artifact, etc.

Fun3. Ability to share existed annotations with other CRIS or with central aggregation systems. E.g. to export data to some central repository which collects existed annotations and provide API to get back all accumulated annotations for the specified resource.

Fun4. Ability to manage taxonomy of motivations and scientific relationships. E.g. in a form of the central repository of controlled semantic vocabularies which were submitted to be used for annotating of research artifacts and relationships between them.

Fun5. Ability to provide notifications for authors of annotations, and annotated resources. Ability for authors of annotated resources to react on information about motivation/character of using their research results, e.g. by providing author's support to improve quality of the annotation, or by improving quality of own material, which was annotated with useful comments, or by protesting against wrong using of their research material, and so on.

Fun6. Ability to collect and process statistics about annotations, its semantics and linked resources. E.g. to build scientometric indicators including aggregated "portrait" of scientific usage for research artifact, publication, author and organization.

This necessary CRIS functionality creates a new phenomenon, which we called the Open (re-)Use since it looks like some evolution of Open Access ideas.

The Open (re-)Use idea 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 ideas and outputs. Some of these relationships become visible as citations in researchers' outputs. 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 a complete picture about the scale and scope of research outputs using and impact.

To respond on this challenge we propose a concept of the Open (re-)Use technology versus traditional paper-based "use", which limitations are well known. The Open (re-)Use technology for the research area should create following new kinds of openness:

1. An openness of results of researchers' manipulation of the materials. It should be clear specified what pieces of the materials were selected by a researcher as artifacts for its further using. From a technical point of view, it should allow scientists to share with the community research outputs in more reusable form than traditional journal article, book, etc. (see Fun2 above).

2. An openness of researchers' motivations to use selected artifacts in producing new scientific knowledge. Only a 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 the artifacts, when there were tries and fails to use it. In that case, the result of researchers' tries and fails in using the artifacts also have to be publicly available (see Fun1).

3. 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 is achieved by creating electronic notification system, which will trace facts of research objects using and will provide information about this for all interested parties (see Fun5).

4. An openness of usage statistics aggregated by a research output, by a researcher (e.g. for all research outputs by this author) and by 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).

Technically, it works as 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 (see Fun6).

4. Requirements for a CRIS-CERIF system

Discussed above the new functionality for a research information system allows us to formulate some additional requirements for designing a CRIS system [10] based on CERIF data model [11].

Req1. CERIF data model should cover new entities like a micro-, a nano-publication, a research artifact and a relationship, a scientific relationship taxonomy and an annotation (see details in "2.6 Requirements" of http://www.openannotation.org/usecases.html), etc.

Req2. CRIS should be integrated with the annotation and research relationship ecosystem to allow both creation and visualization of annotations and linkages for information objects from its content.

Req3. CRIS should be integrated with scientific ontologies ecosystem which is embedded into annotation/linkage technique to help users express explicitly their knowledge, opinions and hypotheses about scientific relationships between research artifacts and so users can visualize in computer-readable form the facts and motivations of using research artifacts for their research process or not using them.

Initial set of rendered scientific relationship classes [8], [12] can include: (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.

Req4. CRIS should support non-traditional forms of research outputs micro-, nano-publication and research artifact, which designed for better reuse. The benefits of depositing research outputs in such new forms are [4]: "(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."

Req5. CRIS should support notification ecosystem to provide a guaranteed awareness of its users about facts of annotating/using their research outputs, including tries/fails data and motivations. The notification system can exist inside or outside CRIS. It should monitors all changes over a set of annotations/linkages existed for information objects of this CRIS and send e-mail notifications to authors of these objects.

The notification service improves scientific circulation and communication because it immediately informs scientists about annotating/using their research outputs. It also improves global research cooperation because researchers can immediately react on how their research outputs were used by the community.

Req6. CRIS should be integrated with statistical ecosystem, which collects and aggregates usage statistics from single CRIS and provides data to the community on how research outputs are used. An approach for designing research artifacts reuse indicators was presented in [12].

5. Conclusion

We believe that open data on how the research data are used to produce a new scientific knowledge will positively influence on current research practice and increase efficiency of the science system at large, and this paper will stimulate discussions on how CRIS-CERIF technology can contribute into this.

Acknowledgements

This research is partly funded by Russian Foundation for Basic Research, grant 12-07-00518-A.

References

1. Neylon C. Architecting the Future of Research Communication: Building the Models and Analytics for an Open Access Future. *PLoS Biol* 2013, 11(10): e1001691. doi:10.1371/journal.pbio.1001691
2. Neylon C. Science publishing: Open access must enable open use. *Nature*, 2012/12/20/print, http://dx.doi.org/10.1038/492348a
3. Groth P, Gibson A, Velterop J. The Anatomy of a Nano-publication. *Information Services and Use* 2010, 30(1/2), http://iospress.metapress.com/content/ftkh21q50t521wm2/
4. Clark T, Ciccarese P, Goble C. Micropublications: a Semantic Model for Claims, Evidence, Arguments and Annotations in Biomedical Communications. 2013, http://arxiv.org/abs/1305.3506
5. 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. http://socionet.ru/publication.xml?h=repec:rus:mqijxk:23
6. Parinov S. The electronic library: using technology to measure and support Open Science. In: Proceedings of the World Library and Information Congress: 76th IFLA General Conference and Assembly, Gothenburg, Sweden, August 10-15, 2010, http://www.ifla.org/files/hq/papers/ifla76/155-parinov-en.pdf
7. 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
8. Parinov S. Towards a semantic segment of a research e–infrastructure: necessary information objects, tools and services. *International Journal of Metadata, Semantics and Ontologies*, 2013, 8(4), 322-331, http://socionet.ru/pub.xml?h=RePEc:rus:mqijxk:30&l=en
9. Parinov S, Kogalovsky M and Lyapunov V. A Challenge of Research Outputs in GL Circuit: From Open Access to Open Use. In Proceedings of the Fifteenth International Conference on Grey Literature: The Grey Audit, A Field Assessment in Grey Literature, 2‑3 December 2013 / compiled by D. Farace and J. Frantzen; GreyNet International, Grey Literature Network Service. Amsterdam: TextRelease, March 2014, http://socionet.ru/pub.xml?h=RePEc:rus:mqijxk:33&l=en
10. Jeffery K. CRIS: Overall Architecture, euroCRIS seminar, 2008, http://www.eurocris.org/Uploads/Web%20pages/seminars/Seminar_2008/Session_5_-_Keith_Jeffery.pdf
11. CERIF data model, http://www.eurocris.org/Index.php?page=CERIFreleases&t=1
12. Parinov S, Kogalovsky M. Semantic Linkages in Research Information Systems as a New Data Source for Scientometric Studies. *Scientometrics*, 2014, v. 98, i. 2, pp. 927-943, http://socionet.ru/pub.xml?h=RePEc:rus:mqijxk:31&l=en