CODATA Data Citation Workshop

Country Report: Russia

Report compiled by: Alexandra Astapenkova

Date: 19.08.2016

Table of Contents

1. Introduction

The Russian Data Citation Workshop was held on 21 July 2016 in Rosa Khutor, Sochi, Russia and was organized in the framework of the international conference “Data Intensive System Analysis for Geohazard Studies” (18–21 July 2016). The workshop was organized by the Geophysical Center of RAS with strong contribution from CODATA. The local organizing committee of the workshop consisted of Alexandra Astapenkova (GC RAS) and Dr. Alena Rybkina with support from other members of the organizing committee of above-mentioned conference. We expected that many of the workshop participants are geoscientists, so the main constraint was to choose the speakers specialists in geosciences that will focuson issues that could be useful, educational and at the same time understandable for wide geo community.

2. Workshop Detail

2.1 Which stakeholders were represented

The attendees of the workshop were the representatives (researchers, leading researchers and chiefs) of different institutions of the Russian Academy of Sciences, specialized mainly in geosciences; representatives of the universities (Russian and foreign – Finnish, Austrian, Italian, Indian) also specializing in geosciences and/or mathematics. Several attendees were involved into the activities related to the data centers and data storages development (such as World Data Centers). The overall number of the participants was 26.

2.2 Workshop content

The workshop included 6 presentations:

1. Data Citation: its importance and the work of CODATA (*Dr. Simon Hodson*, Executive Director, CODATA International Council for Science: Committee on Data for Science and Technology)
2. World Data Center of the All Russian Research Institute of Hydrometeorological Information, (Dr. Aleksandr Sterin, All Russian Research Institute of Hydrometeorological Information)
3. Earth Science DataBase Project – ESDB: geophysical data registration, publication and DOI assignment, (*Alexandra Astapenkova & Dr. Ernest Kedrov*, Researcher & Chief of The Laboratory of Geophysical Data Storage and Dissemination Technology, GC RAS)
4. Information maintenance of the federal state information system "Federal state information system «Unified fund of geological information» (FSIS «UFGI»)", (*Dr. Evgeniya Cheremisina*, Moscow branch FSBI «Rosgeolfond» «VNIIgeosystem»)
5. Soviet and world scholarly works dedicated to rock burst: citation analysis and practices, (*Dr. Andrian Batugin*, MISIS)
6. On uncertainties and dynamics of geophysical data on-line, (*Dr. Vladimir Kossobokov*, Institute of Earthquake Prediction Theory and Mathematical Geophysics)

Valuable contribution to the workshop was made by the presentations of Dr. Hodson, Dr. Sterin and Dr. Kedrov/Ms. Astapenkova.

The Dr. Hodson’s presentation turned out to be the most efficient to the community as the level of awareness to data citation issues was different (e.g. there were some scientists that were not familiar with ORCID).

The presentation, made by Dr. Sterin (World Data Center of the All Russian Research Institute of Hydrometeorological Information), pointed out the main problems and challenges in the field of

geosciences arising in connection with the implementation of the data citation principles. The main problem is the significant difference between “theory and practice” of the data citation in Russia. While the Russian scientific community (researchers, data producers and partly publishers and research performing organizations) is gradually becoming more familiar with the data citation principles and advantages of its implementation, the funding and inspecting organizations (mainly governmental) are planting the usage of quite outdated register systems for data collections and other objects of intellectual property. That is why the existence of the “vicious circle” of non-sharing data is directly and indirectly supported.

The presentation of the Earth Science DataBase Project (ESDB) project, made by Dr. Kedrov and Ms. Astapenkova (GC RAS) shared the clear example of the implementation of the data citation principles and described the mechanism of assigning DOI to geoscience data (in particular, magnetic time series). The presented experience could help geoscientists and researchers, specializing in other fields, see and understand that the technical side of the process is quite simple and achievable. However, the main challenge for the project now is lack of the financial support: the above-mentioned reasons make it quite difficult to find the proper grant as the funding agencies are not aware of the importance of the data citation.

2.3 Key actions/ interventions identified at the workshop

We feel that the key task to be coordinated and taken forward after the workshop is raising the Russian scientific community’s awareness of the data citation principles and promotion of its implementation. We suppose that at the moment there are not enough people in Russia understanding and implementing the data citation principles and at the same time being able to promote the importance if the idea, explain it to funders or other participants of the scholarly communications. That is why we consider that data citation principles education and promotion to be the one of the most important tasks in Russia.

One more major task, identified at the workshop by Dr. Jeffrey Love and Dr. Alexey Gvishiani, is long-term data preservation and other questions related to it, such as the responsibility of the repositories, systems of long lasting data preservation etc. In our opinion, this topic is so important and wide that is even a bit out of the main scope of the workshop. The responsibility of the data publishers is implied when we speak about providing the persistence linking to data (e.g. DOI). As for the data preservation issues, we would like just to add that it is very important to promote the data citation principles among the Russian stakeholders engaged in developing of national standards of data preservation. Moreover, the governmental interest in that field has noticeably grown after the huge loss of paper and digital data destroyed by fire in 2015 in INION (Institute of Scientific Information on Social Sciences of the Russian Academy of Sciences).

3. Main Observations from the Workshop

3.1 What is the policy environment for data citation in Russia?

In our opinion, data producers, data managers and, to a certain extent, the research performing organizations, libraries and publishers in Russia are the most interested in promotion and implementation of data citation principles stakeholders. Funding organizations and research administrators, strongly influencing the opportunities of the scientists to spend more time on developing and implementation of the new principles rather than dealing with the current issues, are not familiar with the benefits of data citation. However, measuring the sciencemetric indices (h-index, impact factor etc.) is becoming more and more popular and influencing: e.g., a researcher, working at a Russian university or an academic institution, could be greatly reduced in extra and even basic payments in case of not satisfying the sciencemetric requirements.

In other words, the real life situation is like that: a researcher or a scientific project manager in Russia has to spend much time on, first of all, finding money on the research and then on writing all kinds of report papers. The situation is further complicated by the obligation of publishing the articles, connected with the projects, in journals with high impact factors. If a project includes e.g. development of a database, the most common way of its acknowledgement (supported financially) is to register outputs of the research as Intellectual Property (IP) Objects. The number of registered IP Objects is a criterion of “good” or “not good” effectivity of science organizations or of individual scientists. Each registered object gets 10-digit unique number, but data base is not necessarily provided for open access (via Internet). The other option in Russia is Joint State Information System of Research Results Registry (EGISU NIOKR) (http://www.rosrid.ru), where separate databases are registered as results of research projects. The unique ID is 12 to 24 digits, the list of metadata is wider than in IP-objects system, but once again no reference to open access in the Internet is provided.

As this practice is supported by the government and required for evaluation of the project results, the scientists (already quite tired of the different evaluations and papers, appearing because of the reform process in Russian Academy of Sciences) have low motivation to invent, employ and implement something new, such as data citation principles. Researchers and data producers feel isolated from the like-minded Russian stakeholders and could be even depressed by the indifference of the authorities (in particular, The Federal Agency for Scientific Organizations (FASO Russia), managing nowadays the Russian Academy of Sciences) to the needs of the researchers.

We would like to underline that talkingabout the funding organizations in Russia, especially in case of the fundamental science, we very often imply the governmental structures, such as above-mentioned FASO or the Ministry of Education and Science of Russia because, e.g. each institute of the Russian Academy of Sciences is obliged to fulfill its own so-called “government plan”, and in case of failure working with the other grants becomes impossible. These are the factors hindering fast and wide distribution of the data citation principles in Russia (we use the geoscientific community as an example but believe that there are the same issues in other fields of Russian science).

If we try to discourse of the policy that would most effectively advance science through encouraging data citation, we would like to express several ideas. First of all, in our opinion, education and detailed explanation the basic principles of data citation can be very useful not only for the Russian researchers, but also for the governmental representatives and other stakeholders, such as librarians or archivists. We believe that growing awareness of data citation principles and ideas of both researchers and scientific authorities will one day result in a consensus among all the stakeholders. Consequently, activities, aimed at involvement of all kinds of stakeholders into examination and distribution of the data citation principles and ideas, seem to be the most promising and effective. E.g., active Russian researchers and data managers could encourage the organizers of the conferences on corresponding topics to invite foreign experts, specializing in data citation (such as CODATA representatives) for giving presentations, moderating workshops etc. Secondly, the support of the local projects and activities related to data citation, is very important. We would like to underline that guidance and consultative support is as important as the financial support. Moreover, we suppose that appearance of detailed but clear (and, maybe, translated to Russian) publications (articles, blogposts etc.) devoted to data citation, supported by concrete examples and cases, could also help the increase of data citation awareness.

3.2 What infrastructure is available to support data citation?

According to our knowledge, data centers and data archives in Russia are usually managed and operated independently by scientific institutions or universities and there is no agreed common standard of its work. In the sphere of geosciences there are good examples of data centers , such as World Data Centers that are accepted as the Regular Members of the World Data System (WDS) (http://www.wdcb.ru/index.html). The above-mentioned Earth Science DataBase Project (ESDB), initiated by Geophysical Center RAS, includes but

is not limited to improvement of the Russian WDC infrastructure with regard to geophysical data registration and DOI assignment.

Repositories, locally developed by universities or institutes in Russia often provide assignment the persistent identifiers, including DOI, but no special attention is usually paid to data publishing and citation. We would also like to mention here Elibrary and Russian Science Citation Index (http://elibrary.ru/) – the largest and, probably, the most influential domestic Russian system, that is available to all researchers and could become an "infrastructural bridge" between the cited data and scientists that want to cite data. Being the DOI-friendly system, Russian Science Citation Index is a bibliographic database of scientific publications in Russian that accumulates more than 2 million publications of Russian authors, as well as information about citing these publications from more than 2000 Russian journals, scientific reports, conference proceedings etc. We believe that Elibrary's specialists could be interested in implementation of the data citation principles, although, our practical knowledge shows that this organization is often overloaded by large amount of tasks related to development and support of its own database, bibliometric evaluation of journals etc.

Last "infrastructural contact" that is not directly related to data centers or archives, but could help to involve into the process of data citation principles' distribution the large groups of Russian stakeholders, such as scientific publishers and academic librarians, is Russian Association of Science Editors and Publishers (ASEP, http://rasep.ru/). The President of the Association Olga Kirillova is doing a lot of organizational and educational work devoted to raising the standards of the Russian scientific publishing, so we suppose that data citation principles and standards could be promoted in Russia with the help of ASEP and, especially, annual international conference "World-Class Scientific Publication" (http://conf.neicon.ru/index.php/science/domestic0516), organized by ASEP in Russia.

Persistent identifier systems, that are available to all researchers in Russia and are governmentally supported, were mentioned and described in chapter 3.1. These are:

1) The Open Register system, provided by Rospatent – the Federal Service for Intellectual Property managed by the Ministry of Economic Development of the Russian Federation (http://www.rupto.ru/about?lang=en, http://www1.fips.ru/wps/portal/Registers/). Databases and other objects of intellectual property, registered in that system, get a 10-digit persistent identifier supported by the metadata description. Search on the system's website is available online only by the serial number of the database and could not be called user-friendly. Moreover, the metadata descriptions of the databases contain no links to websites and have only the names of its owners.
2) Joint State Information System of Research Results Registry (EGISU NIOKR) (http://www.rosrid.ru). Separate databases are registered there as results of research projects. The unique ID contains from 12 to 24 digits, the list of metadata is wider than in Rospatent system, but it is quite difficult to find a database there, as the system is designed for registration and search of the scientific projects that does not always contain databases.

In our opinion, both systems, mentioned above, are widely used in Russian scientific practice, mainly because of the obligation to use it and not because of its convenience or efficiency.

3.3 What are current attitudes to data citation?

To a greater or less extent, we have already described the factors, significantly influencing the current attitude to data citation in Russia in Chapters 3.1–3.2. Speaking in general, not so many researchers and other stakeholders in Russia are familiar with data citation. Researchers, using and citing data in their routine scientific work, deal with it in their own way. The secondary use of data still is currently an option. As no governmental reward system for scientists makes provision for the publication of data sets as a specific scientific output, citing of data, using the persistent identifiers (besides plain text descriptions or links to websites), is rare.

The participants of the workshop, that are familiar with the data citation principles, supported the Joint Declaration of Data Citation Principles developed by Data Citation Synthesis Group (FORCE11) and "Out of Cite, Out of Mind: The Current State of Practice, Policy, and Technology for the Citation of Data", elaborated by CODATA-ICSTI Task Group on Data Citation Standards and Practices, and noticed that "it is difficult to add something". However, the main issue for the Russian stakeholders is the practical application of the mentioned-above principles.

3.4 Benefits and challenges in implementing data citation policies and practice (including economic and financial considerations)

We would like to mention the benefits of implementing data citation policies and practice, identified during the workshop, as challenges were mainly detailed in previous chapters. The benefit for data producers, managers and depositors seems to be quite clear: e.g., if data, produced and verified by an observatory, is cited, it means that it works "for a reason" and number of citations, in its turn, can help to prove to the funding agencies (such as FASO) observatory's being in demand in fundamental science. One more identified benefit related to researchers that can use and cite data and then be sure in the persistence of the links and data (of course, if the data producer and/or depositor works in accordance with the data citation principles).

The big identified challenges are (besides the general situation in Russian science) data preservation and data quality control, mentioned in Chapter 2.3.

As for the proposed solutions, the main (and, maybe, the only) accessible way of changing the "data citation climate" in Russia is education and promotion of the data citation principles and organization of the workshops and seminars devoted to the popularization of the topic. The probable solutions were also described in the last part of Chapter 3.1.

4. In Conclusion

4.1 What is the role of the research funding and policy community in implementing data citation policies and practices?

When we speak about promotion and sharing the information regarding the data citation principles, it seems to us that each stakeholder, i.e. a researcher, a data producer etc. plays an important role in the process of changing the attitude to data citation. As governmental structures play great role in Russian scientific life, it is also very important to explain to them why it is important and how it can be improved. That is the reason why good practical examples or, at least, thoroughly elaborated papers (especially in Russian) about the development and implementation of the data citation principles, could be supportive and useful. The initiatives, primarily based on a bottom-up approach, usually seem to be more effective in Russian scientific community as Russian scientific decision-makers and government representatives (including foundations supporting R&D) are usually more welcome to support the clearly elaborated projects and initiatives. Thus the research community should thoroughly prepare and explain why implementation of the data citation principles is such an important initiative.

4.2 Summary of points of convergence and specific plans for implementation

As we have already mentioned in Chapter 3.4, the agreement reached during the workshop is to continue sharing information about data citation principles and promotion of its implementation. Probably, the next educational step would be activities devoted to data citation, organized as a part of the Regional CODATA Eurasia conference, which will be held in September-October 2017.

4.3 Reflections on the workshop

If we expect to arrange the next workshopwe would like to make it more educational and to invite, ideally, as many Russian stakeholders as possible. From the first stages of preparation we were limited by those people who had an opportunity and could afford to come to Sochi that is why there were less participants than a real quantity of geoscientists, interested in the topic. However, the feedback that we received is still helpful, illustrative, and reflecting the situation at least in Russian geoscientific community (and, to some extent, it could be even extrapolated to other branches of science). One more important aspect that we would probably change or at least take into account is the working language of the workshop. The organizers of the Sochi conference refused to provide the simultaneous interpreting (because most of the conference participants spoke English and for financial reasons), but we suppose that in case of discussion of cutting edge topics it is very important to get all the small details and to involve into discussion the stakeholders that are not so fluent in English. Providing more time for the workshop would also definitely add to the facilitation of the communication between the participants.

We consider the educational component to be the aspect of the workshop that worked really well. The opportunity to ask questions and consult an international expert (Dr. Hodson) was also supported by the participants.

Finally, we would like to thank all the colleagues that inspired and helped us to organize the Russian Data Citation Workshop. There is a Russian proverb “The first pancake is always lumpy”, meaning “it's the first step that is troublesome”. We hope that our pancake was not too lumpy, was quite tasty and contained some calories. We also believe that this experience will help us to promote and implement data citation in Russia and organize new workshops and other activities that will benefit both Russian scientific community and the world scientific community.

Attachment 1: Attendance Register/ Overview

1) Dr. Aleksandr Sterin	All Russian Research Institute of Hydrometeorological Information
2) Dr. Andrian Batugin	National University of Science and Technology MISiS
3) Dr. Anni Reissell	International Institute for Applied Systems Analysis
4) Dr. Antonella Peresan	Department of Mathematics and Geosciences, University of Trieste
5) Dr. Ernest Kedrov	Geophysical Center RAS
6) Dr. Evgeniya Cheremisina	Moscow branch FSBI «Rosgeolfond» «VNIIgeosystem»
7) Dr. Jeffrey Love	USGS Geomagnetism Program
8) Dr. Hanna Lappalainen	University of Helsinki
9) Dr. Jyrki Lappalainen	University of Helsinki
10) Dr. Kusumita Arora	CSIR-National Geophysical Research Institute, Hyderabad, India
11) Dr. Nina Frolova	Institute of Environmental Geoscience, RAS
12) Dr. Seyed Naser Hashemi	School of Earth Sciences, Damghan University, Damghan, IRAN
13) Dr. Simon Hodson	CODATA, ICSU Committee on Data for Science and Technology
14) Dr. Tatiana Khromova	Institute of Geography RAS
15) Dr. Vladimir Kossobokov	International Institute of Earthquake Prediction Theory and Mathematical Geophysics RAS
16) Dr. Vladimir Kostsov	St.Petersburg State University
17) Dr. Vyacheslav Gusyakov	Institute of Computational Mathematics and Mathematical Geophysics SB RAS
18) Mr. Artem Smirnov	Lomonosov Moscow State University, Faculty of Geology
19) Ms. Alexandra Astapenkova	Geophysical Center RAS
20) Prof. Alexey Gvishiani	Geophysical Center RAS
21) Prof. Marat Zakhidov	Bio Science Incorporated

22) Prof. Mikhhail Gokhberg	Schmidt Institute of Physics of the Earth RAS
23) Prof. Petr Martyshko	Institute of Geophysics, UB of RAS
24) Prof. Petr Shebalin	International Institute of Earthquake Prediction Theory and Mathematical Geophysics RAS
25) Prof. Valery Vernikovsky	Novosibirsk State University and Institute Petroleum Geology and Geophysics SB RAS
26) Prof. Viktoria Kulikova	Institute of Geology Karelian Research Centre Russian Academy of Sciences

Attachment 2: Workshop Programme

Russian Data Citation Workshop – Agenda
21 July 2016

09:00 Welcoming remarks and introduction, *Dr. Simon Hodson,* Executive Director, CODATA International Council for Science: Committee on Data for Science and Technology

09:10 Data Citation: its importance and the work of CODATA, *Dr. Simon Hodson,* Executive Director, CODATA International Council for Science: Committee on Data for Science and Technology

09:40 Questions and General discussion

09:50 World Data Center of the All Russian Research Institute of Hydrometeorological Information, *Dr. Aleksandr Sterin,* All Russian Research Institute of Hydrometeorological Information (2 min for questions)

10:00 Earth Science DataBase Project – ESDB: geophysical data registration, publication and DOI assignment, *Dr. Ernest Kedrov*, Chief of The Laboratory of Geophysical Data Storage and Dissemination Technology, GC RAS (2 min for questions)

10:10 Information maintenance of the federal state information system "Federal state information system «Unified fund of geological information» (FSIS «UFGI»)", *Dr. Evgeniya Cheremisina*, Moscow branch FSBI «Rosgeolfond» «VNIIgeosystem» (2 min for questions)

10:20 Soviet and world scholarly works dedicated to rock burst: citation analysis and practices, *Dr. Andrian Batugin*, MISIS (2 min for questions)

10:30 On uncertainties and dynamics of geophysical data on-line, *Dr. Vladimir Kossobokov*, Institute of Earthquake Prediction Theory and Mathematical Geophysics (2 min for questions)

10:40 Panel discussions:
«Data Citation Culture implementation in Russia: a need, a long term perspective or an achievable goal»
«Data – Management, Sharing and Services: Current & Future Role of Data Stakeholders»

11:20 Summary of points of convergence and next steps, *Dr. Simon Hodson*

11:30 Meeting adjourns

11:30–12:00 Coffee break