



CRIS 2014

## Particular experience in design and implementation of a Current Research Information System in Russia: national specificity

V.A. Zelepukhina, T.S. Danilova, A.S. Burmistrov, Yu.Yu. Tarasevich \*

*Astrakhan State University, 20a Tatishchev Street, Astrakhan, 414056, Russia*

---

### Abstract

We present our experience to implement a Current Research Information System in the Astrakhan State University, Russia. The CRIS covers research information such as publications, funding, projects, and patents. Education, training, professional activity, and expertise are also within the scope of the CRIS. The metadata of research output are stored as a single record, regardless of the number of contributors and their belonging to different divisions. Data linkage associates a research output with contributors and their departments. A faculty fills own profile and links the metadata with co-authors. The more active and responsible users employ the CRIS, the more complete and reliable is information. Some metadata are imported from external data sources, in particular from Russian Science Citation Index. CRIS can generate Statistical Reports, CV's, Publication Lists, ratings, etc. When implementing the system the main problems are not technical but organizational. The bureaucratic obstacles hinder the exchange of information between different sources. The managerial staff does not demonstrate the real interest in getting actual, complete and accurate information about the research outputs. Some of the faculty has rather little computer skills and low interest in filling the database.

© 2014 The Authors. Published by Elsevier B.V.

Peer-review under responsibility of euroCRIS.

*Keywords:* Current Research Information System; Database, Fuzzy Double, Metadata, Research Output, Interacting with External Sources

---

### 1. CRIS of Astrakhan State University (Russia)

Since 2011, our team makes efforts to design, implement, and put into operation a CRIS <http://science.aspu.ru> in Astrakhan State University. 834 faculties are registered in the CRIS at April 1, 2014. The CRIS contains metadata about 5595 publications, 563 patents, 92 funding, 19 contracts, 137 theses (dissertations), and 171 awards.

---

\* Corresponding author. Tel.: +7-8512-610925; fax: +4-8512-494157.

*E-mail address:* [tarasevich@aspu.ru](mailto:tarasevich@aspu.ru)

The main principle of implementation is metadata of research output are stored as a single record, regardless of the number of contributors (co-authors) and their belonging to different divisions (departments or labs)<sup>1</sup>. Mechanism of data linkage associates a research output with contributors and their departments.

Our experience allows formulating some common problems that arise in the implementation of CRIS in scientific organizations in Russia. Some of these problems are specific to Russia, associated in particular with the transition from the Soviet system of education and research to the European. Additional problems arise from high level of bureaucracy.

### 1.1. Structure of the CRIS

CRIS of the Astrakhan State University includes<sup>1</sup>:

- database that stores information about the participants of scientific activity (employees and departments);
- database that stores metadata of research outputs;
- tools for analysis and identification of links between participants of scientific activity and the research outputs;
- tools to eliminate redundancy in the database by identifying fuzzy duplicates in the metadata of research outputs;
- tools to interacting with external sources of scientific and scientometric information.

Interface of the CRIS includes several types of pages.

1. Pages of divisions such as departments, laboratories, institutes, research centers, etc.
2. Personal pages of faculty.
3. Pages of scientific schools and scientific directions.

The last item probably needs explanation. The research groups (so-called scientific schools and directions) might differ from departments and laboratories. Annual Statistical Reports are requested from departments as well from official confirmed schools and directions while structures of the reports are completely different. At present, there are 119 divisions, 8 scientific schools, and 23 scientific directions in Astrakhan State University.

Pages of divisions and research groups include several main thematic sections.

1. Common information about a division or a group.
2. Staff.
3. Publications.
4. Patents.
5. Funding.
6. Contracts.
7. Awards.
8. Defenses of the theses.
9. Activities (seminars, conferences, etc.).
10. Statistics.

Personal page of a person additionally contains:

1. some personal information, position, undergraduate and graduate education, Post Graduate education and defenses of theses including habilitation, trainings, academic degree and title,
2. links to profiles in scientific networks as Academia, Mendeley, ResearchGate, and in external Databases as Russian Science Citation Index, Google Scholar, ResearcherID, Scopus Author Evaluator, Orcid, arxiv.org, PubMed.

Faculty profiles include the following sections

1. publications;
2. funding;
3. patents;
4. awards;
5. activities including
  - supervising of students and Post Doctoral researchers,
  - defenses of theses of the supervised students,

- public reviewing of the theses,
- courses taught,
- supervising of the Graduate and Post Graduate programs,
- editorial positions,
- membership in the Dissertation Councils,
- expertise,
- membership in the Organizing/Advisory/Program Committees of the conferences, exhibitions, competitions, etc.,
- participation in the conferences, exhibitions, competitions, etc.,
- membership in professional societies.

## 1.2. Services

Additional possibilities of the CRIS are

1. A messaging system.
2. Additional options for the administrator (notifications of new entries in the database, notification of potential duplication of information, notification of possible supplies of scientific output to a department or a person, etc.).
3. Publication ranking<sup>5</sup> and other analytics.
4. Constructor for lists of publications.
5. CV Constructor.
6. Notification about potential co-authoring of a research output added by other user.
7. Generation of statistical reports of departments, scientific schools and directions.
8. Search and filtering of the information.
9. Automatic assignment of specific tags to journal articles (e.g., the journal belongs to so called list of Supreme Certification Commission).
10. Searching for fuzzy doubles in metadata of publications.

## 2. Input and verification of the metadata

### 2.1. Ways to fill the database

A faculty should fill own profile with metadata of research output and link the metadata with co-authors. The more active and responsible users employ the CRIS, the more complete and reliable information is stored in the database.

At present the CRIS has not an access to some local databases of the Astrakhan State University such as Human recourses, PhD Study, etc. because of bureaucratic obstacles.

Additional problem is the subscription to Scopus and Web of Science. The price is too high for a regional university.

Nevertheless, it is possible to import some metadata from external databases and sources<sup>2</sup> (Fig. 1).

1. The CRIS is in position to recognize and extract same useful information from theses in MS Word format. Unfortunately, theses in PDF format are not always generated in a proper way, so a program is hardly able to recognize a Cyrillic text and extract the metadata.
2. Scientometric information from Russian Science Citation Index is downloading using free API.
3. Citations can be downloaded automatically in BibTeX or in format of one of two All-Russian State Standards 7.1–2003 ‘Bibliographical Description’ and 7.0.5–2008 ‘Citation’. Publication metadata can be downloaded by DOI through CrossRef.
4. Information about patents is downloaded from the local Database of the Astrakhan State University.
5. A part of information about professional expertise (membership in the dissertation councils) is imported from the database <http://science-expert.ru>.

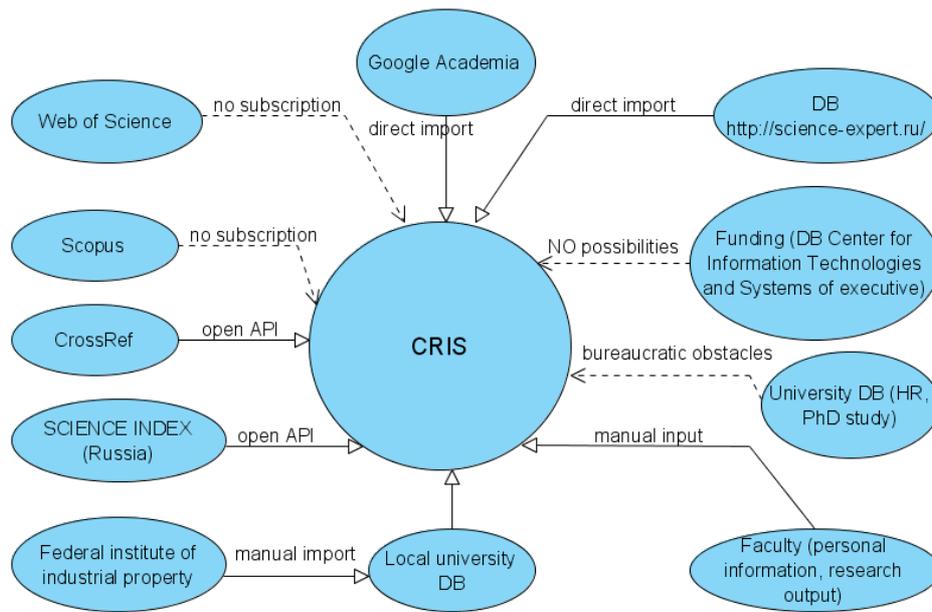


Fig. 1. Different sources and ways to put metadata of research output in CRIS of Astrakhan State University. Solid arrows indicate the implemented data streams, dashed arrows show possible but not used at present streams.

## 2.2. How to eliminate typical errors in metadata

The above mentioned problems lead to the fact that a substantial part of the information can not be automatically imported from reliable sources. Large amount of information is downloaded manually. This information is subject to further verification and validation to the doubles. Partly these problems can be resolved by automatic data linkage and duplicate search. For the manually entered information, there is a problem to check the links between research output and contributors, to find and remove duplicates, to supplement the missing information.

Due to the manual input of metadata, some typical errors arise frequently<sup>3</sup>:

- the data is incomplete;
- typos in the output title, contributor's name and other attributes;
- usage of abbreviations and acronyms in the title, publisher name, journal title and etc.;
- the order of contributors (co-authors) is incorrect;
- volume, issue or pages are incorrect or missed.

The errors may produce the fuzzy doubles. We designed the program which detects fuzzy doubles of research output metadata<sup>4</sup>. The program works as follows:

- Generating stemmed forms for the string attributes of metadata.
- Fuzzy comparing of the string attributes using n-grams.
- Clustering of the potential doubles using Euclidean distance.

A particular problem is data linkage of information obtained from external sources and the contributors. We designed a special program for data linkage between research output co-authors and employees and its departments (labs). The program uses the methods of fuzzy comparison and some transliteration rules for data processing. This approach is chosen because the Cyrillic names can be transliterated in English in different ways. Some times one name can be transliterated in many ways, e.g. Scopus author ID 6603962335: Tepyĭ, Tepyĭ, Tepy, Tepyĭ,

TEPLYI, Teply, Tyopliy, Telyi. Possible variants of the name of the employee are generated and stored in the CRIS.

### 3. Main problems

There are several national peculiarities affecting the development of CRIS. First of all, one should mention small national experience in design and using CRIS. Even the leading university in Russia, Lomonosov Moscow State University, finished testing and began full operation of CRIS <http://istina.msu.ru/> in 2012. One of the oldest CRIS in Russia is used in Joint Institute for Nuclear Research <https://pin.jinr.ru/pin/pin> since 2008.

1. Low level of information technologies usage during exchange of information between research institutes, universities and organizations engaged in accounting of the research output. Basically, information is entered manually; there is no exchange of files in a standard format.
2. Lack of possibilities for automated integration with other national databases (e.g., Center for Information Technologies and Systems of Executive) increases the laboriousness of the database filling;
3. Only small amount of universities and research institutes has a subscription to Scopus and Web of Science; as a consequence receiving reliable scientometric information is difficult.
4. There is an alternative citation index for publications in Russian, i.e. Russian Science Citation Index; scientometric information of Russian Science Citation Index is required for statistical reports in Russia.
5. There is a large variety of indicators and their instability in the statistical reports of the research organizations.
6. There are several specific characteristics in metadata of articles as ‘journal belongs to so called list of Supreme Certification Commission’, ‘journal is indexed by Russian Science Citation Index’, ‘translated journal’).
7. Only a small part of publications of faculty is indexed by Scopus and Web of Science.
8. Only a small part of articles in Russian has DOI, therefore, the automatic import of metadata from the reliable sources (e.g. CrossRef or directly from the publisher) is difficult.
9. Hybrid (Soviet-Bologna) system of education and academic degrees produces additional inconvenience.
10. There is a huge amount of different classification schemes.
11. The formal structure of the organization may not coincide with the real scientific groups (scientific schools and directions). This creates additional difficulties in preparation of statistical reports and analysis of scientific activity, because reports on the structural units (departments and labs) and scientific schools and directions should be generated separately.
12. Bureaucratic obstacles difficult to obtain information from both external and local sources.  
Additional problems arise during operation of CRIS.
  1. Low interest of faculty in filling the database (lack of financial incentives).
  2. Some of the faculty has rather little computer skills.
  3. Nobody reads any user manual and system messages. Hence, step-by-step interface is required to input any information.
  4. The managerial staff does not demonstrate the real interest in getting actual, complete and accurate information about the intellectual activity.

### 4. Conclusion

Among the problems encountered in the implementation of CRIS, we can highlight the technical, organizational and financial aspects. The most complex of these are organizational problems.

The lack of real interest of the managerial staff in the use of information technology to obtain actual, complete, and accurate information about the scientific activity is a major obstacle when using CRIS.

Stream of information ‘employees–database’ is very unreliable: the information comes irregularly, incomplete and unreliable. If uncertainties and incompleteness can be partially eliminated by technical methods, the problem of irregularity can be solved by administrative methods only.

The obtained experience assists to create a test version of an academic network <http://dropsnet.itmmse.ru/> based on the similar ontology as the CRIS.

## Acknowledgements

The reported study was partially supported by Russian Foundation for Humanities, research project No. 12-03-12000 and by Russian Foundation for Basic Research, research project No. 12-07-31145.

## References

1. Zelepukhina VA, Tarasevich YY. The concept of data-processing system for the collection and analysis of scientific and scientometric information in the organization. *Informatization of Education and Science* 2013; **2(18)**:133-144.
2. Zelepukhina VA. The integration of bibliographic databases with the external web services. *Informatization of Education and Science* 2014; **1(21)**:179-189.
3. Zelepukhina VA. The problems of the accuracy and objectivity of the information stored in the scientific Web 2.0 community. *PRIKASPIYSKIY ZHURNAL: Upravlenie i Vysokie Tekhnologii (CASPIAN JOURNAL: Management and High Technologies)* 2013, **4(24)**:157-164.
4. Umarov AS, Popova NV, Zelepukhina VA. Some aspects of the development of information systems for the collection and storage of scientific and scientometric information. *PRIKASPIYSKIY ZHURNAL: Upravlenie i Vysokie Tekhnologii (CASPIAN JOURNAL: Management and High Technologies)*. 2013, **3(23)**:111-118.
5. Tarasevich YY. Scientometrical Information and its interpretation. *Informatization of Education and Science* 2014; **2(22)**:141-148.