Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 00 (2016) 000–000

Procedia Computer Science

www.elsevier.com/locate/procedia

13th International Conference on Current Research Information Systems, CRIS2016, 9-11 June 2016, Scotland, UK

Limitations of CRIS in assessing the progress of increasing research output in UrFU

Mark Akoev[a,b]*, Olga Leyba[a] , Lev Golitsyn[b]

[a] *Scimetrics Lab of Ural Federal University, Lenina 51 134, Ekaterinburg 620000, Russia*
[b] *Integrirovannye Sistemy LLC, Presnensky val 19b1, Moscow 123557, Russia*

Abstract

Using CRIS metrics to monitor and stimulate research activity is a widely accepted practice. We implemented CRIS Pure in UrFU and in this paper we summarize limitations of CRIS we met and outline the approaches proposed to solve them.

Keywords: Cohort Analysis; ArchiMate; Information model of Research Process; Development of Research Group; world-class university; Research information management systems (RIMS)

1. Introduction

Ural State Federal University is facing an ambitious goal: become a world-class university and secure a position in a top 400 of world university rankings[1]. This is the main purpose of the 5-100 Project[2], a government program the university is participating in, that is aimed to maximize the competitive position of a group of leading Russian universities in the global research and education market. Due to ratings being heavily based on scientific contribution, the main objective of UrFU now, in accordance with the 5-100 program, is to rapidly and drastically increase volume and quality of research output.

The history of Ural Federal University (UrFU) begins in 1920. By the data available for 2014, the year used to analyze research output, the University employed 3200 academic and research staff, 26000 full-time students were

* Corresponding author. Tel.: +7-904-98-82823.
E-mail address: m.a.akoev@urfu.ru

studying in UrFU and 900 of them were in PhD programs. 4000 papers were published in 2014 affiliated with UrFU, but only 29% of them are indexed in Scopus and Web of Science. 71% of papers were published in local Russian peer-reviewed journals.

The position university currently holds is represented in Fig. 1, compared to universities chosen as benchmarks for the 5-100 program. Performance metrics are chosen for this graphics to indicate the main problem we are dealing with - an extremely, as compared to benchmarks, low proportion of active authors to research and academic stuff. While other shown metrics, like publications per author or even the number of publications itself, are in line with other universities, the proportion of affiliated authors with indexed papers to the number of staff is just 0.6, with only 1952 unique authors presented in Scopus. As the picture indicates, a corresponding coefficient for foreign universities is higher than 1, which suggests high scientific output of PhD and MS students.

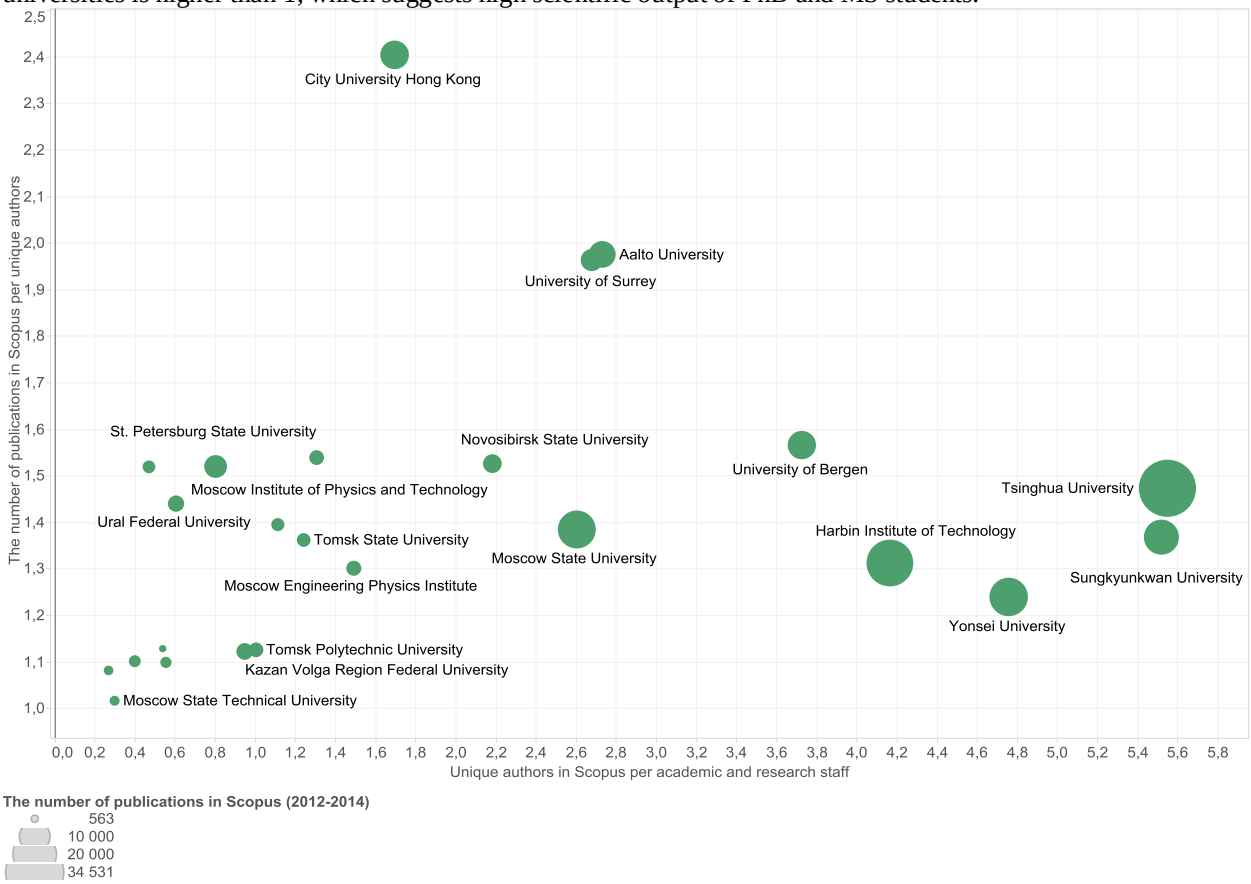

Fig. 1. Comparison chart of author publication activity. Source: SciVal 2012-2014. Information about academic and research staff from The Report Russian Ministry of Education and Science 2013 and InCites 2012.

While 5-00 Project provides participating universities with additional financing (approx. 6-15 mln Euro per year), the amount doesn't allow for inviting of already formed research groups. Therefore, the chosen approach to increase the performance is to enhance the development of a new and existing research groups from staff and students of the university.

2. Introduction

The first approach to stimulate the increase in research output of the staff that was tested in UrFU was to award authors of papers indexed in Scopus and Web of Science Core Collection. An internal program was established through 2011-2014 to reward a bonus for each paper indexed. The bonus was substantial as compared with a regular salary, estimated at around 260% of average monthly FTE academic staff salary.

As a result, the percentage of staff that has one or more papers published grew, but only from 17% in 2011 to 23% in 2014, stopping completely in 2015. The rise of 10% can be viewed as almost insignificant in the perspective of metrics shown above. Overall, the results of bonus program were found not encouraging, and the program was discontinued.

Comparison of 23% active staff with 60% unique authors shows that a substantial amount of activity may be attributed to PhD students. Unique authors contribution shows not only staff and students, but also other affiliations, and more accurate data currently isn't available. Nevertheless, if we estimate PhD activity by figures shown above, it's safe to assume that students are much more productive than staff.

An analysis of bonus program awards data has shown that people who participated the most and gained the most benefit was the people who had published successfully before the program has started. This finding suggested that the gap to publishing in an indexed journal is severe enough that people will not overcome it themselves just because being provided with an incentive to do so. The analysis also revealed that the distribution of successful authors by departments is severely skewed and did not show a tendency to even during the program. New participants in the bonus program tended to emerge from departments where the share of academic staff with published papers already was at 40% and higher. Therefore, it was deemed necessary to facilitate the scientific activity among academic staff and students and provide means to transfer research related skills[3] from existing research groups to the rest of academic staff in an interdisciplinary and even in an interdepartmental way.

3. CRIS Pure as a way to see the whole scientific picture

In order to find perspective ways to improve scientific activity and to assess the effects of actions taken it was required to acquire a better, more detailed view of a research activity inside the university. It was decided that data on all papers, including those in local peer-reviewed journals, needs to be acquired and analyzed, as well as data about all grants that academic staff are participating in.

The hypothesis was that papers in local journals could be used as indicator of a promising scientific activity that could subsequently or with aid result in indexed publications. Also, it was important to have accurate affiliation data that would again encompass all scientific contributions in order to get sight of collaborations network and identify promising lines to direct attention to.

CRIS Pure by Elsevier was chosen as a tool to acquire and organize data. The implementation started in July 2014. We have met and successfully overcame a number of obstacles in the course of the project, such as adaptation to Cyrillic and special aspects of local employment, academic and education structure.

As no sources of local publications data are available for import, the initial scheme of data input was to engage authors themselves with a subsequent approval of designated coordinators. The approval step here is crucial as the data is to be used as a primary source for the staff performance evaluation. Unfortunately, the approval scheme that would be supported by Pure and found reliable by university authorities is not found yet, which impedes the engagement of authors. This obstacle challenges the goal of obtaining all the data, as at a university scale it is highly questionable to have the data fully maintained by designated personnel.

Another objective of Pure implementation was for its Portal part to serve as a tool for self-promotion of researchers and establishing new collaborations. But, as the usage of a new system by academic and research staff has shown, people who were eager to take advantage of this part of the system were the people already quite successful in their academic careers. As the target stated above was to increase the productivity of a less productive or unproductive stuff, the effect of Pure as a tool to promote scientific communication in such a situation is questionable.

4. System engineering approach on development of research groups

When our primary objective is to have all the academic staff involved in research, it is important to find ways to engage those who are currently not participating in it. It is also important to work with active researchers too, investigating our options to improve citation impact. Here the impact is not the end goal by itself, but an indicator of participation in a broad scientific communication, which in turn is a predictor of a high-demanded research.

Implementation of projects targeted towards these two goals is complicated by a long feedback loop. We can't expect a newly formed research group to publish sooner than a couple of years and citations will occur even later. Same feedback timeframe can be expected for effects of actions taken to improve the impact of existing groups if we measure only the impact itself. This makes such projects a high-risk investment, especially keeping in mind that in order to achieve our objectives we need a significant amount of new research groups to be developed and involved in the mainstream scientific communication.

We made an attempt to model the research group working process in UrFU environment as an enterprise architecture using Archimate modeling language[4] and put together scientific practices, university facilities, and stakeholder concerns. The resulting model, if verified to be accurate, could be used as a tool to predict future results by assessing intermediate benchmarks. Transferable skills analysis from[3] was used as a basis for the model. The resulting model and the approach to it was discussed in an INCOSE RUS bimonthly meeting[5].

The model introduces four subsequent levels of a research group, each characterized by a list of practices established in a group, such as an ability to organize regular seminars, continuous student recruiting, alignment of educational programs content with current scientific research needs and so forth. A study to verify that each level is associated with a certain level of research output quality is underway.

If the study shows that maturity level of a group is indeed a strong predictor of its scientific success, we will have a tool to estimate the university's future progress by a system dynamic model based on the group lifecycle. A fine-grained assessment on levels of all existing scientific groups would allow to plan targeted actions on improving the practices lacking where it is needed and allocate resources accordingly. However, to conduct such analysis a large variety of data to assess maturity levels of groups will be required and CRIS Pure currently is not suited for collecting this, as it was not intended to be used this way. A dedicated Research Information Management System (RIMS)[6] could be employed to collect and aggregate the required data.

One example of data that could be useful is information on publications reference groups, which characterize the research area of a new or existing research group. Reference groups can be outlined as a part of the dedicated development plan for particular groups as a direction for collaborations broadening. They could also serve as a standard of research presentation needed to publish at a WoS/Scopus level.

As a part of UrFU strategic research development plan, Thomson Reuters provided an analysis of qualitative and quantitative indicators of UrFU publications compared to three Russian and five foreign universities used as benchmarks. This study confirmed materials science to be the UrFU main research area, providing 31% of all publications from the 2010-2014 period. However, citation impact of materials science publications is the lowest in UrFU and even lower than the average impact for Russian publications. Upon further investigation, it appeared that the reason for such low impact is not the quality of the research itself but rather the lack of proper presentation. Over 90% of UrFU publications in materials science were classified into 10 topics, and for each topic a reference group was suggested[7]. Five topics were identified as a top-priority by a university counsel, and a further analysis was conducted to propose a list of key researchers, institutions and journals for each topic. Resulting recommendations were presented to principal investigators.

However, before such collaboration evidence as publications and citations no traceable collaboration links appear in Pure and thus Pure as a CRIS does not allow to monitor how these recommendations are followed through. A system that would collect data on communications, such as visits or internships, would be useful for this purpose. This information could also be used to verify and refine our view of reference groups, especially when it does not include actual successful international collaborations. In order to keep this view in use and up-to-date it is also needed to represent and maintain it as structured data linked to Web of Science rather than as a document. Yet, communication data should not be used as a measured objective of the research group to avoid manipulation and serve as a meaningful predictor of future improvement.

For newly formed research groups we need to assess a basic activity that can be supported on a large scale without huge investments, and reading is a perfect candidate here. We can analyze EZproxy[8] logs analysis to gauge the amount of requests for research-related resources, as it is used to access them from outside of UrFU campus both by staff and students. A steady increase of requests amount could be an early predictor of effect for projects aiming to engage new people in research. Data on downloads supplied by publishers could be used in the same way. As EZproxy data is linked to personal accounts, a study could be conducted to verify if requests activity is a predictor of research activity.

5. Another ways to assess intermediate progress

Indicators assessing research groups practices, such as those suggested in section 4, should be corroborated by indicators verifying the progress at the university scale. Being able to trace predicting indicators to research output metrics would allow to assess the meaningful effect size of research facilitation projects undertaken, which would be extremely helpful.

The most used research output metric is the number of publications in Scopus or WoS CC with at least one author listed as affiliated with UrFU. This number shows steady growth. The persistent growth rate of the number of WoS CC publications also indicates an increase in the number of high-quality publications (Fig. 2). However, we can not identify which of actions undertaken caused the increase. Also, it should be taken into account that metrics of scientific results are prone to manipulations and various biases[9].

As it was shown in section 1 (Fig. 1), author productivity in UrFU is already on a good level, so in order to increase the university output we need to increase the numbers of authors without causing a drop in author productivity, or, even more importantly, research quality. Hence the university level indicators should be able to demonstrate how the number of active authors and their productivity changes over time, otherwise a growing detrimental effects of various stimulating programs and employment of assessing tools may remain hidden by summary statistics for a very long time and become insusceptible to simpler corrective measures. University level indicators should answer the following main questions: how quickly retiring authors are being replaced by new authors; how productive these new authors are, and whether new authors continue to work in UrFU as they grow.

The similar problem of assessing progress of a company exists in startups[10]. In this case, a research group development could be treated as a startup, and publications as novel products in the target market. Most measures and actions used in research group development are similar to those common in promotions of novel or modified products. In both cases, it is important to conduct several controlled experiments with different designs of actions and activities and to be able to reliably measure the impact of those actions and activities on participants (scientists in the case of research groups). Cohort Analysis is a reliable measurement tool for this task, and was proven effective[11].

Cohort Analysis suggests that year of first publication activities by new scientists can be used in the measurement of the impact of development actions. In the absence of specific indicators (described in 4), we can not measure the exact moment when a new scientist joins the research group, but we still can use a date of the first publication as the cohort indicator for them. Additional analysis and further corrections for the author's cohort could be done using CRIS data (Pure data in the case of UrFU). Cohort Analysis can be applied on the university, the department and the subject area levels, which makes it extremely useful.

Fig. 2 shows the structure of authors' cohorts construction. All authors who published a paper in the first year of the analysed period are treated as new authors (including those who were active before). Every following year new authors are added. New authors are PhD students or researchers who previously did not affiliate with UrFU or published in local journals only. As authors can skip a year between publications, and PhD students do not necessarily continue to work in UrFU after graduation, we can't expect all new authors to reappear in next years. However, it is important to see that new cohorts does not deplete completely over a period of several years, as well as assess not just the number of authors in a cohort but also their productivity.

Cohort Analysis as applied to authors is illustrated in the Table 1 below.

Table 1. Cohort structure and dynamics

Year	2010	2011	2012
Authors of cohort 2012			New authors who published in 2012: • 15 (PhD students 2012) • 20 (New Researchers, may be with history)

Authors of cohort 2011		New authors who published in 2011: • 10 (PhD students 2011) • 8 (New Researchers, may be with history)	Authors who published in 2011: • 4 (PhD students 2011) • 7 (New Researchers, may be with history)
Authors of cohort 2010	• 10 (PhD students 2010) • 10 (Researchers)	Authors who published in 2010: • 5 (Researchers, ex PhD students) • 7 (Researchers)	Authors who published in 2010: • 4 (Researchers, ex PhD students) • 9 (Researchers)

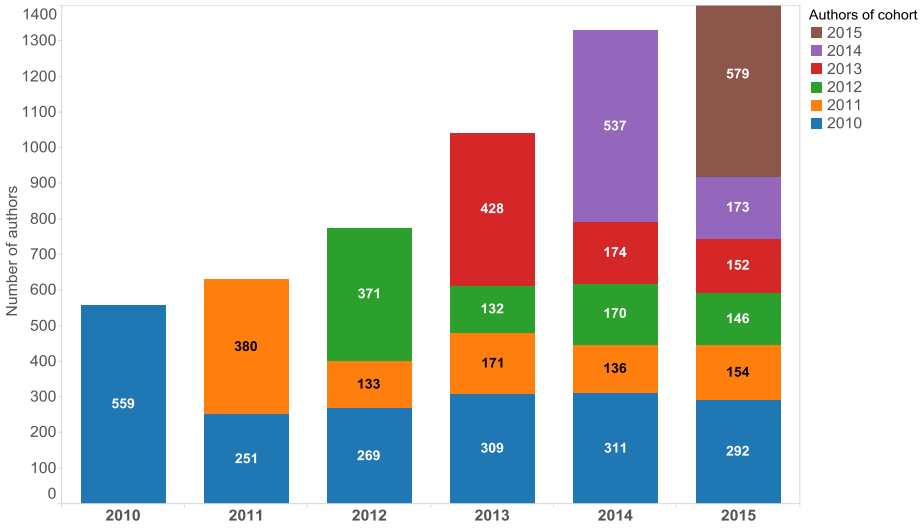

Fig. 2. Cohort analysis of UrFU authors for years 2010-2014

In order to counter such manipulation of authors numbers' metric as adding co-authors to papers, we need to monitor author productivity along with author numbers. A illustration of how Cohort Analysis for publications can be designed is presented in the Table 2. Papers are aggregated by the year of publication and assigned to a cohort spanning over the years when authors started to publish. For example, a paper where three authors started in 2010, 2011 and 2012 would be assigned to the 2010-2012 cohort.

Cohort Analysis of publications is complicated due to most of them having more than one author, and natural co-author composition being a collaboration of new researchers with more seasoned ones. In fact, a measure of how often new authors are collaborating with successful ones can be a predictor of scientific practices transfer and quality level preservation. A further complication to monitoring groups of authors in cohorts is a possible migration of authors between them. For example, an author that started in 2011 can appear in four out of five possible cohorts of year 2012: 2010-2011, 2010-2012, 2010-2011, 2011-2012, but not in the 2012-2012. Further along, even if he was in the 2010-2012 cohort of year 2012, he can move to the 2011-2011 cohort of year 2013. Nevertheless, publication-based cohorts on a university level can still demonstrate to maintain distinctive behaviour and thus be a valid target of assessment, but this hypothesis needs further research.

Table 2. Cohort Analysis of publications dynamics

Publications of authors from cohort	2010	2011	2012
2012-2012			20
2011-2012			5
2011-2011		8	14

2010-2012			10
2010-2011		6	4
2010-2010	10	9	12

Conclusion

CRIS allows for assessment of research performance of those of the academic staff who are already working at a level of publishing papers in indexed journals.

A model of research group maturity was proposed to aid facilitation and monitoring of the development of new research groups. Extending the CRIS capabilities to those of RIMS is needed in order to study the validity of the model and to employ it for action planning.

A cohort analysis method based on CRIS Pure data is proposed to validate the long-lasting effect of actions taken to increase the university's research output.

Acknowledgements

The work is supported by Ministry of Education and Science of the Russian Federation, research project No. 14.579.21.0117 (id RFMEFI57915X0117).

References

1. ACTION PLAN for implementation of the Competitiveness Enhancement Program of the Federal State-Owned Autonomous Educational Establishment of Higher Vocational Education "Ural Federal University Named After the First President of Russia B.N. Yeltsin" Among the world's leading science and education centres for 2013 – 2020. 2013. 188 p.
2. The Program of Enhancement of Competitiveness of Russian Universities (Project 5-100) [Internet]. [cited 2016 Feb 20]. Available from: http://5top100.com/
3. Transferable Skills Training for Researchers / OECD. Paris: Organisation for Economic Co-operation and Development, 2012. 151 p.
4. ArchiMate [Internet]. [cited 2016 Feb 20]. Available from: http://www.opengroup.org/subjectareas/enterprise/archimate
5. [Architecture of science and engineering work in a university] [Internet]. [cited 2016 Feb 20]. Available from: http://incose-ru.livejournal.com/54648.html
6. Research information management systems – a new service category? [Internet]. 2014. [cited 2015 Nov 18]. Available from: http://orweblog.oclc.org/Research-information-management-systems-a-new-service-category/
7. Research project: development of strategic business plan. Phase 2: Materials science - targeted analysis. Report for Ural Federal University and the Ural Branch of the Russian Academy of Sciences. 13th November 2014. Prepared by Thomson Reuters IP&Sciences Consultancy Group.
8. EZproxy Overview [Internet]. 2015 [cited 2016 Mar 22]. Available from: https://www.oclc.org/ezproxy.en.html
9. Butler L. Explaining Australia's increased share of ISI publications—the effects of a funding formula based on publication counts. Research Policy. 2003 Jan;32(1):143–55.
10. Eric Ries The Lean Startup: How Today's Entrepreneurs Use Continuous Innovation to Create Radically Successful Businesses. 2011.
11. Cohort analysis. In: Wikipedia, the free encyclopedia [Internet]. 2015 [cited 2016 Feb 20]. Available from: https://en.wikipedia.org/w/index.php?title=Cohort_analysis&oldid=690818506