

Evaluation methodology for monitoring the usage of open science platforms



Document prepared by Petar Čolović

WP4 - Building Capacities for Implementation of Open Science Principles

C4 - Evaluation methodology for monitoring the usage of open science platforms and the performance and impact of research at Serbian universities

2020

The European Commission support for the production of this document does not constitute an endorsement of the contents which reflects the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein

One of the crucial tasks in implementing the open science principles in Serbia is enabling the sustainability of the established infrastructure. Besides ensuring the full technical functionality of the institutional repositories, the aspect of monitoring and evaluation of the usage of open science platforms, as well as of the research impact, as stated in the final work task of the Work Package 4 (WP4).

Under the roof term “evaluation methodology” we imply a set of indicators for monitoring the use of institutional repositories, as well as monitoring the impact of research within Serbian academia, as stated in the WP4. In the process of development of such methodology, our focus was primarily on assembling a reliable set of indicators with the capacity to both accurately show and diagnose the current state of affairs, as well as with capacity to detect changes and serve as a prediction tool. The aforementioned features, along with the principle of simplicity, were derived from the Leiden manifesto (translated into the Serbian language by Smederevac&Pajić, available at <http://beopen.uns.ac.rs/events.php?event=217296fcd8e51ec608690fdf5082f269>). A set of documents relevant in this field (please see the reference list below), was used as the basis for our methodology. Initially we drafted a model according to the recommendations by Blumel, Dietze, Heller, Jaschke, and Mehlberg (2014). The draft can be seen at http://beopen.uns.ac.rs/documents/e8504ec7b56311dae310cbbe2368c21a/Presentations_.BEOPEN_WP4_PColovic_23.03.2018.pdf. The model comprises the following elements:

1. Indicators for monitoring the usage of open science platforms:
 - a. Number of item downloads
 - b. Number of item views
 - c. Number of open publications (further divisible into numbers of pre-prints and post-prints)
2. Indicators for evaluating the impact of research at Serbian universities
 - a. Traditional research measures, as recommended by the University of Leiden Centre for Science and Technology Studies (CWTS; <http://www.cwtsbv.nl/monitoring-evaluation>) including, but not limited to, mean normalized citation score, general scientific output, collaboration measures)
 - b. Alternative set of measures under the umbrella term “Altmetrics”, including, but not limited to, data obtained from blogs, Twitter, Mendeley, social networks, media.
 - c. The indicators are set to be applied to the following “units of analysis”:
 - i. Individual authors
 - ii. Research groups
 - iii. Institutions (at different levels of academic hierarchy)
 - iv. Publications available at the institutional repositories.

Alongside the two sets of indicators, and according to the Leiden manifesto and the Report of the European Commission Expert Group on Altmetrics, the two groups of indicators can be accompanied by (complementary) qualitative methods, including peer review.

The fully developed institutional platforms based on DSpace – CRIS contain the data appropriate for all the above mentioned methodological and analytic procedures. Publication-related data comprise information on authors, the relevant metadata, publication abstract, availability of the full text (along with the full-text file if available), SCOPUS-based citation number, the number of page views, number of downloads, and the access to the Google scholar item and the accompanying information. The author-related data useful for the monitoring and evaluation comprises the information contained in the Author Profile, such as the information on number of publications, as well as the author’s collaboration network.

While a number of basic yet informative statistics can be found in the repository itself, including its homepage, more advanced analyses can be performed by extracting the data in a standard “wide” data format (Figure 1). A standard set of variables within such datasets includes, but is not limited to, the data on *id*, *collection*, *authors*, *doi*, *availability of full text*, and *grant*.



Figure 1. Data extracted from the repository. Taken from http://beopen.uns.ac.rs/documents/e8504ec7b56311dae310cbbe2368c21a/Presentations_WP4_PColovic_23.03.2018.pdf. Image blurred for the protection of the privacy of the data.

To illustrate the possible use of the data for monitoring purposes, a simulated dataset was used (R Core Team, 2020; Venables & Ripley, 2002), utilizing the information available at the homepage of the University of Novi Sad repository. Thus a simulated dataset contained the assumed 3503 author entries, 13349 publications, 10000 journal publications, 2933 conference presentations, the assumed 522 full texts including 374 preprints (estimated according to SCOPUS, www.scopus.com). The data is not the accurate representation of the current state of affairs, but a simulated dataset appropriate for illustration purposes only.

Table 1. Example: Descriptive Statistics for authors

		jour	confer	other	open	itations	iews	ownloads
	Vali	nal arcticles	ence papers	research outputs	research outputs			
d		3503	3503	3503	3503	503	503	503
sing	Mis	0	0	0	0			0
n	Mea	7.278	4.981	3.760	3.915	.503	5.897	.861
Deviation	Std.	1.046	1.048	1.032	1.062	.038	.039	.046
imum	Min	3.000	1.000	0.000	0.000	.000	2.000	.000
ximum	Ma	11.000	8.000	7.000	7.000	.000	9.000	.000

The techniques can be applied in the same or similar fashion to the indicators of alternative metrics.

For the follow-up studies, for the purposes of monitoring the change in scientific output, repeated measures analyses can be performed by extracting the data from the repository, matching it

by author and conducting the analyses using simple or advanced analytic procedures for dependent data.

The set of techniques constituting a methodology for monitoring the usage of open science platforms is capable of producing the information relevant for evaluation, assessment and planning of the use of institutional repositories. It is possible that the present set of techniques may be complemented by additional procedures along with the increase of the available data.

Blümel, I., Dietze, S., Heller, L., Jäschke, R., & Mehlberg, M. (2014). The quest for research information. *Procedia Computer Science* 00, 000–000. Available at https://dspacecris.eurocris.org/bitstream/11366/194/3/30_Blumel_et_al_CRIS2014_Rome.pdf.pdf

Hicks, D., Wouters, P., Waltman, L., De Rijcke, S., & Rafols, I. (2015). Bibliometrics: the Leiden Manifesto for research metrics. *Nature*, 520(7548), 429-431.

R Core Team (2020). R: A language and environment for statistical computing. R foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.

JASP Team (2019). JASP (Version 0.11.1)[Computer software].

Venables WN, Ripley BD (2002). *Modern Applied Statistics with S*, Fourth edition. Springer, New York. ISBN 0-387-95457-0, <http://www.stats.ox.ac.uk/pub/MASS4>.

The Leiden Manifesto for Research metrics [in Serbian]. <http://www.leidenmanifesto.org/translations.html>;
[http://beopen.uns.ac.rs/documents/217296fcd8e51ec608690fdf5082f269/The%20Leiden%20Manifesto%20for%20research%20metrics%20\[in%20Serbian\].pdf](http://beopen.uns.ac.rs/documents/217296fcd8e51ec608690fdf5082f269/The%20Leiden%20Manifesto%20for%20research%20metrics%20[in%20Serbian].pdf)