Building a solid foundation for better monitoring: Utilizing Persistent Identifiers in Serbian National CRIS to advance metadata quality and authorship association

Authors

Vladimir Otašević[2, 1] (ORCiD: 0000-0002-9408-3454), Biljana Kosanović[1] (ORCiD: 0000-0003-0543-4767), Dušan Stijović[1]

[1]University of Belgrade, Computer Center
[2]University of Belgrade, School of Electrical Engineering

Abstract

The eNauka platform is a national Current Research Information System (CRIS) in Serbia. The platform utilizes Persistent Identifiers (PIDs) to enhance research metadata quality, streamline authorship association, and integrate with citation databases. Automated authorship association saves time for end-users, simplifying research management processes. With over 200 registered scientific organizations and 20,000 active researchers, eNauka serves as a vital portal for observing the national scientific community. Its commitment to standardized metadata collection methodologies ensures data consistency and accuracy across diverse scientific domains, facilitating informed decision-making and further innovation in research.

Keywords

eNauka platform, Persistent Identifiers (PIDs), Metadata quality, Research impact assessment, Automated authorship association

The eNauka platform is a national CRIS designed to keep track of scientific research in Serbia. Specifically, it collects and stores research output from across different fields and provides a way to assess the performance of individual researchers and organizations based on their contributions. It's continuously improving to meet the changing needs of the research community in Serbia. Overall, eNauka makes it easier to manage and evaluate scientific work in the country. It helps evaluate researchers and research organizations by digitizing important processes, like:

1) Scientific accreditation
2) Career Progression in Academia

For these procedures to succeed and to accurately depict the state of science within scientific organizations and among researchers, it's crucial to have accurate data. The eNauka platform aims to simplify the collecting of metadata about publications, saving time for researchers and administrators from each scientific organization (known as NIO editors). This simplification helps ensure that the data input process is efficient and reliable, enabling a comprehensive overview of scientific activities in various domains.

When possible, the development team behind the eNauka platform strives to introduce a mechanism for automatically supplementing missing metadata in scientific records. However, acknowledging the

limitations of full automation, certain features are designed as hybrids. This means that while the platform can automatically fetch metadata, it requires verification by the NIO editor to guarantee accuracy and completeness.

The eNauka platform encompasses three primary entities: Results, Researchers, and Scientific Research Organizations. These entities are interconnected, with changes made to one instance inevitably influencing the results presented by the others.

Figure 1 showing relation between entities

The success of the eNauka platform critically depends on its ability to effectively address practical challenges that impact the quality of the results presented within the system. Two primary challenges include:

1) Establishing and Maintaining Entity Relations: One key challenge revolves around establishing and maintaining relationships between different entities within the platform, such as results, researchers, and scientific organizations. Simplifying the process of establishing and maintaining these relationships is essential for facilitating seamless data management and analysis within the platform.
2) Foundation for Monitoring and Evaluating Researchers and Organizations: Another crucial challenge involves laying the groundwork for monitoring and potentially evaluating researchers and scientific organizations based on their contributions and activities. This requires implementing comprehensive frameworks and metrics for assessing research productivity, impact, and collaboration, while also ensuring fairness and transparency in the evaluation process.

Addressing these practical challenges is fundamental for the eNauka platform to achieve its goal of delivering accurate and reliable insights into the research landscape. By developing effective solutions to streamline entity relations and establishing frameworks for monitoring researchers and organizations, the platform can significantly enhance its utility and impact in supporting better decision-making.

Persistent Identifiers serve as a critical component of every entity within the eNauka platform. By utilizing collected PIDs, eNauka can uniquely identify each instance of an entity. PIDs are gathered in various ways across the eNauka platform. Where it is possible, they are obtained from external services and national sources via implemented APIs. However, in certain scenarios, manual intervention by NIO editors is necessary to add PIDs. Even in these instances, the eNauka platform's development team aims to optimize user efficiency by providing suitable forms for the addition of PIDs, thereby saving time for end-users. Errors can occur during PID entry, which is why the eNauka platform strives to offer automatic semantic and syntactic checks for PIDs whenever feasible. Furthermore, certain PIDs, such as DOIs, undergo validity checks by attempting to resolve the full path to which the DOI leads.

Metadata Quality

Metadata about publications in eNauka is obtained through two methods. The first method involves scheduled harvesting, where organizations with registered OAI-PMH servers meeting minimal technical requirements for harvesting are included for periodic retrieval (Lagoze et al. 2002; Jackson et al. 2008; Walk and Baker 2018). This approach is particularly suitable for organizations with established infrastructure (Ivanović et al. 2017). By maintaining their existing deposition methods, organizations can seamlessly continue depositing records, while eNauka automatically retrieves metadata from their servers (Otasevic, Kosanovic, and Stijovic 2023). As an alternative, researchers can retrieve their results from external services such as CrossRef, Scopus, PubMed, and others.

In certain cases, these records may contain only minimal metadata. Therefore, the eNauka platform endeavors to enhance them with PIDs whenever feasible. By doing so, these records will incorporate additional information that was not present in the original metadata source.

On the other side, the eNauka platform stands out by its ability to fetch citations from diverse databases, setting it apart from other previously used platforms. National legislation places significant emphasis not just on scientific productivity but also on citations within specific databases (Министарство просвете, науке и технолошког развоја 2020). Also, platform eNauka is trying to be as much as possible in correlation with nation platform for open science (Ministry of Education, Science and Technological Development of the Republic of Serbia 2018). Given its national-level data collection on scientific productivity, the platform aims to standardize methodologies for gathering metadata regardless of the scientific field from which a particular scientific result originates.

Enriching available records in eNauka with PIDs extends beyond enhancing descriptive and identifying metadata. By incorporating PIDs into records, the eNauka platform can offer more sophisticated features for end users. Given that eNauka serves over 200 registered scientific organizations with approximately 20,000 active researchers whose scientific outputs are accessible through multiple sources harvested by the platform, implementing a deduplication mechanism becomes imperative. PIDs serve as one of the technical methods used by the platform to detect existing results, thus preventing the accumulation of duplicates.

Authorship Association

The collection of well-described scientific results is only valuable if there are clear and valid connections between entities. To ensure data quality, the eNauka platform utilizes PIDs to accurately identify authors of specific scientific results.

The accurate linkage between researchers and scientific results is indirectly correlated with the achievements of a particular institution. This is because scientific productivity, measured by the publication of scientific results, is heavily influenced by the individuals associated with the institution. Consequently, evaluating the citation impact of an institution relies on the impact of specific results published and achieved by researchers affiliated with that institution.

Introducing ORCiD into eNauka has transformed how researchers engage with their work globally. ORCiD, short for Open Researcher and Contributor ID, is a tool that empowers researchers to boost their visibility

and credibility in the scientific community (Arunachalam and Madhan 2016; Haak et al. 2012). With each researcher assigned a unique and persistent digital identifier, ORCiD ensures accurate attribution of contributions, eliminating confusion caused by similar names and ensuring proper recognition of their work (Otasević and Kosanović 2022).

Moreover, ORCiD seamlessly integrates with various research systems and platforms. Researchers or NIO editors can link ORCiD identifiers to publications, datasets, monographs, and other scientific results. By using ORCiD, eNauka helps researchers to manage their identity, become more visible, and contribute to the scientific community (Kosanovic 2017).

The eNauka platform primarily utilizes ORCiD as the main PID for identifying authorship on a paper. To establish this connection, two criteria must be met: the ORCiD must be listed in the metadata of the harvested scientific records in eNauka, and the ORCiD must be assigned to a researcher registered in the national registry of researchers. Once these criteria are fulfilled, every scientific result associated with this particular ORCiD, whether through OAI-PMH or another workflow, will be automatically attributed to the researcher linked to that ORCiD. This automated association of authorship saves time for end-users, eliminating the need for manual checks on authorship.

Conclusion

The platform eNauka plays a vital role in facilitating research management and scholarly communication in Serbia. Through its use of PIDs, integration with citation databases, and commitment to metadata enrichment, the platform ensures the accuracy, visibility, and better quality of research outputs. Through its integration with established institutional infrastructure, eNauka enhances their further development.

References

Arunachalam, S., and M. Madhan. 2016. “Adopting ORCID as a Unique Identifier Will Benefit All Involved in Scholarly Communication.” *The National Medical Journal of India* 29 (4). http://dst.sciencecentral.in/16/.

Haak, Laurel L., Martin Fenner, Laura Paglione, Ed Pentz, and Howard Ratner. 2012. “ORCID: A System to Uniquely Identify Researchers.” *Learned Publishing* 25 (4): 259–64. https://doi.org/10.1087/20120404.

Ivanović, Dragan, Dušan Surla, Miroslav Trajanović, Dragan Misić, and Zora Konjović. 2017. “Towards the Information System for Research Programmes of the Ministry of Education, Science and Technological Development of the Republic of Serbia.” *Procedia Computer Science* 106: 122–29. https://doi.org/10.1016/j.procs.2017.03.044.

Jackson, Amy S., Myung-Ja Han, Kurt Groetsch, Megan Mustafoff, and Timothy W. Cole. 2008. “Dublin Core Metadata Harvested Through OAI-PMH.” *Journal of Library Metadata* 8 (1): 5–21. https://doi.org/10.1300/J517v08n01_02.

Kosanovic, Biljana. 2017. “Mala video škola za ORCID.” *RCUB Media portal* (blog). October 19, 2017. http://media.rcub.bg.ac.rs/?p=5973.

Lagoze, Carl, Herbert Van de Sompel, Michael Nelson, and Simeon Warner. 2002. “Open Archives Initiative - Protocol for Metadata Harvesting - v.2.0.” June 14, 2002. http://www.openarchives.org/OAI/2.0/openarchivesprotocol.htm.

Ministry of Education, Science and Technological Development of the Republic of Serbia. 2018. “(Translation) Open Science Platform.” MESTD. http://www.open.ac.rs/images/doc/Open-Science-Policy-Serbia.pdf.

Otasević, Vladimir, and Biljana Kosanović. 2022. “Resolving Authorship Using the ORCIR Identifier.” In *28th Scientific and Buisiness Conference YU INFO*.

Otasevic, Vladimir, Biljana Kosanovic, and Dusan Stijovic. 2023. “Establishing Metadata Exchanging Procedures for Increasing Computer Interoperability/Compatibility and Metadata Quality.” Presented at the Open Science Fair 2024, Madrid, Spain, October 2. https://doi.org/10.5281/zenodo.8397964.

Walk, Paul, and Tom Baker. 2018. “The Use Of Persistent Identifiers In Dublin Core Metadata.” In . https://github.com/dcmi/pids_in_dc/blob/master/report/Report_DC2018_PID_Session_20180911.md.

Министарство просвете, науке и технолошког развоја. 2020. “Pravilnik o Sticanju Istraživačkih i Naučnih Zvanja: 159/2020-82, 14/2023-51.” December 25, 2020. https://www.pravno-informacioni-sistem.rs/SlGlasnikPortal/eli/rep/sgrs/ministarstva/pravilnik/2020/159/18/reg.