Utilizing Existing Data Resources and Open Infrastructures: The Implementation of CRIS in Serbia (eNauka)

Biljana Kosanović, Vladimir Otašević, Dušan Stijović

University of Belgrade, Computer Center

Introduction

The paper presents the basic entities and functionalities of eNauka - a publicly visible information system and portal created to provide an overview of scientific research activities in Serbia. eNauka includes: 206 Research Performing Organization (RPO), 19.980 researchers and more than 700.000 publications.

In Serbia, there have been several attempts in the last twenty years to establish a national CRIS, but none have succeeded in satisfying all needs, especially in terms of the completeness and quality of data, or end-user support. Some of the systems mentioned are: IRIS, cris-NS (Ivanović, et al. 2017), KNR-Vojvodina (Provincial Secretariat for Higher Education and Scientific Research 2011), Dositej (MPNTR 2017), E-CRIS.SR (IZUM 2023) (Albahari 2017), RIS (MPNTR 2018), BeOpen (Smederevac, et al. 2020). All aforementioned systems have been supported in some way by the ministries responsible for science (financially or organizationally), indicating clearly that CRIS is an essential tool for data-driven decision making.

For these reasons, at the end of 2021, the Office for Information Technology and Electronic Administration issued a public call for the construction of a Registry of Researchers in Serbia, based on which subject of this paper, information portal eNauka (NITRA 2023) was developed.

System Description

eNauka is a publicly accessible portal for monitoring the scientific performance of researchers and institutions in the Republic of Serbia, which is aligned with international standards and practices to establish interoperability for data transfer. It is based on DSpace-CRIS (DSpace-CRIS. Version 6.3. 2021) solution, and behind eNauka is a flexible data model that enables the collection and management of research data and information typical for CRIS systems, defining entities and attributes through their reciprocal relationships. The system allows for the management of research profiles (e.g., name variants, ORCID, identifiers from other systems in Serbia, Scopus ID, etc.), the use of standard and various persistent identifiers (e.g., DOI, COBISS-Id, ScopusId, etc.), linking researchers to institutions, and so on. Such data models are called CERIF (Common European Research Information Format (EuroCRIS 2023)), and the data model applied in eNauka is developed according to this standard.

eNauka is comprised of two closely interconnected parts:

- An administrative part that ensures the registration of accredited RPOs in the RPO Register and the registration of researchers in the Researcher Register, and
- A publicly visible registry of results, which includes basic data about researchers and RPOs

Data in eNauka - PIDs

In order to increase the interoperability of the system and utilize positive work-flows in the exchange of scientific information, data about RPOs, researchers, and results have been enriched with persistent

identifiers, or PIDs. This enables connections with existing information systems in the country (such as E-CRIS.SR, KNR), as well as verification of the accuracy of result data (such as DOI, PubMed, etc.).

Data in eNauka - Sources

The system initially loaded all results (publications; cc. 410.000) that existed in the previous RIS (Research Information System). The conversion was performed in May 2022, regardless of the fact that the data were not deduplicated at that time (a significant number of papers appeared multiple times), and the majority were unverified.

Immediately following the data conversion from the RIS, data from other sources in the RPOs were collected, provided they met the necessary condition of having an OAI/PMH server. This protocol for exchange was chosen as it is most commonly used for exchanging scientific information. These include:

- institutional repositories. In June 2022, 45 RPOs maintain a repository/information system that met this condition (Kosanović, et al. 2019). By October 2023, this number had increased to 116 (https://time.graphics/line/314977). About half of the institutions in this increase were not newly established repositories but had already had a repository or information system where OAI/PMH was implemented. Since data entry, verification, and supplementation in repositories are mostly done by librarians, the data collected in this route are not additionally verified for accuracy in eNauka.
- The NaRDuS (National Repository of Doctoral Dissertations, (MPNTR 2015)) is also regularly harvested as a consolidated source of all doctoral dissertations defended in Serbia.
- data from COBISS began to be regularly harvested (weekly) from May, 2023. This source is valuable because COBISS covers the entire publishing production of the Republic of Serbia according to the Law on Mandatory Copy.
- In September 2023, data transfer from the Research Register of Vojvodina (KNR) was enabled, as this system has established the good practice of researchers in Vojvodina entering data about their results into KNR since 2012.
- In October 2023, the harvesting of data from Naši u WoS (National Library of Serbia, Belgrade 2023) (Timotijević, Kosanović and Vasiljević 2013) is started.

The collection of additional result data is initiated by the researcher, and the correctness of the metadata is verified by RPO-editors. This way, the responsibility for the completeness of the data lies with the researchers, while the quality of the metadata lies with the RPO-editors.

Researchers registered on eNauka could log in exclusively using their ORCID iDs, using the same credentials for authentication as for editing their ORCID profiles. This was made possible due to NITRA (NITRA 2023) becoming a regular member of the ORCID organization, and huge campaign aimed at helping users register and maintain their ORCID profiles (RCUB 2017).

Data about publications in eNauka

Each publication in eNauka is equipped and visible with all its PIDs, sources where it is available, and its citation in relevant sources. Citations can only be tracked for publications that have recorded PIDs. Among all PIDs, the DOI stands out because:

- Other PIDs (ScopusID, PMID, etc.) are automatically assigned based on it. A special application has been developed to add other PIDs without the need for manual input.
- It alerts to the correctness of the data in terms of the number of authors in the publication in eNauka and the number of authors assigned by the publisher when assigning the DOI and depositing metadata in CrossRef.
- It indicates the existence of duplicate records.
- It establishes a direct link to the landing page at the publisher's or aggregator's website, i.e., the one that assigned the DOI.
- It alerts to incorrect DOI assignment.
- It collects citations from open sources (OpenCitations. Data repository n.d.).
- It checks the openness of publications for full-text access (Impactstory - Unpaywall n.d.).
- It determines the mention of publications on social media (Altmetric. Service n.d.).

Citation data are periodically retrieved in regular cycles from available sources using PIDs for each individual publication. This further means that if a publication does not have a PID in a service where citation tracking is conducted, then there is no data on citation. Additionally, citation tracking is only conducted in sources that have clearly (and transparently) displayed procedures for deduplication, sources from which they gather data, selection of sources for referencing, etc.

Training and end-user support

Quality of metadata about publications in eNauka is the responsibility of RPO-editors, and special attention has been given to their education. Online (Zoom) training sessions for RPO-editors were held in 5 sessions, and the entire recorded material is available to everyone at https://enauka.gov.rs/regres. Additionally, detailed instructions for work were provided.

Researcher training was conducted exclusively through (1) recorded video tutorials and (2) instructions. In both cases, tutorials of different lengths were prepared (e.g., "Shorter is not possible" or "eNauka on one page"), but there are (rare) instances where even that was not viewed or read. By mid-February 2024, over 11.500 researchers had logged into the system at least once.

End-user support was immediately established, but the number of people engaged in this activity changed. This is a necessary and binding service that must function because experiences from previously established systems indicate that this was a crucial drawback. Portals have good quality to operate 24/7, and users are inclined to utilize that quality. This means that a large number of eNauka users, especially researchers, log in during nights and weekends. Perhaps they do not expect a response during those periods, but if they receive one, they are certainly pleasantly surprised.

Conclusion and further plans

Establishing a national CRIS is not a simple task, as if it were, it would have already been established. Time will tell whether NITRA's decision to finance the development of a national CRIS instead of subscribing to existing foreign commercial solutions was good or not. The result will only be measurable after a minimum of three years, but what is already visible is:

- Human capacity in RPO has been improved in terms of understanding the flows of scientific information.

- The number of repositories has significantly increased.
- Researchers' trust that they will not have to constantly enter their data from scratch has increased.
- Citations are collected from various sources, creating conditions for creating different indicators of success.
- Broad public accessibility contributes to data accuracy.
- Easier and faster rectification of identified deficiencies, and adaptation to domestic regulatory acts.
- Data collection from all domestic sources has been enabled.

Certainly, researchers and RPO-editors play a key role in eNauka. If they accept the system without major resistance, the chance of success increases. However, the decision on the survival of eNauka is ultimately made by NITRA, in several ways: (1) providing institutional user support, (2) providing IT support and development in the same manner, (3) adapting regulations and rules, and (4) planning annual funds for eNauka.

References

Albahari, Biljana. 2017. "Javno dostupne bibliografije istraživača Srbije – stanje i perspektive." *Čitalište : naučni časopis za teoriju i praksu bibliotekarstva* no. 30, pp. 80-94. doi: 10.19090/cit.2017.30.80-94 .

n.d. *Altmetric. Service.* https://www.altmetric.com/.

2021. *DSpace-CRIS. Version 6.3.* https://www.4science.it/dspace-cris.

EuroCRIS. 2023. *CERIF in Brief.* https://eurocris.org/eurocris_archive/cerifsupport.org/cerif-in-brief/index.html.

Impactstory - Unpaywall. n.d. *"Unpaywall" Software.* https://unpaywall.org/.

Ivanović, Dragan , Dušan Surla, Miroslav Trajanović, Dragan Mišić, 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: 122-129.

IZUM. 2023. *Informacioni sistem o istraživačkoj delatnosti u Srbiji.*

Kosanović, Biljana, Milica Ševkušić, Vasilije Rajović, and Nenad Popović. 2019. "Setting the scene for a sustainable national repository network in Serbia." *Open Science Fair 2019, Porto, Portugal.* doi:10.5281/zenodo.3509971.

MPNTR. 2015. *National Repository of Dissertations in Serbia.* https://nardus.mpn.gov.rs/.

—. 2018. *Registar istraživača Srbije.* https://ris2.mpn.gov.rs/.

—. 2017. *Uspostavljen jedinstven registar ustanova i zaposlenih sa nastavno-naučnim, naučnim, istraživačkim i stručnim zvanjima u DOSITEJ-u.*

National Library of Serbia, Belgrade. 2023. *Naši u WoS [Elektronski izvor] : članci naših autora u servisu Web of Science od 2000. godine.* 10 10. https://kobson.nb.rs/nauka_u_srbiji/nasi_u_wos.3.html.

NITRA. 2023. *eNauka.* https://nitra.gov.rs/cir/nauka/enauka.

—. 2023. *Srbija je članica ORCID-a.* https://nitra.gov.rs/lat/ministarstvo/vesti/srbija-je-clanica-orcid-a.

2024. *OpenCitations. Data repository.* https://opencitations.net/.

Provincial Secretariat for Higher Education and Scientific Research. 2011. *Karton naučnog radnika Vojvodine.*

RCUB. 2017. *Mala video škola ORCID-a.* http://media.rcub.bg.ac.rs/?p=5973.

Smederevac, Snežana , Dejan Pajić, Sanja Radovanović, Silvia Gilezan, Petar Čolović, and Branko Milosavljević. 2020. "Funkcionalnosti platforme DSpace-CRIS." In *Otvorena nauka: praksa i perspektive*. Novi Sad: Univerzitet u Novom Sadu.

Timotijević, Tatjana, Biljana Kosanović, and Milan Vasiljević. 2013. "KoBSON: A good news delivery service." *INFORUM 2013: 19th Annual Conference on Professional Information Resources Prague, May 21-22.*