

Nikola Šćepančević

Faculty of Mathematics, University of Belgrade

INSTITUTIONAL REPOSITORY OF THE FACULTY OF MATHEMATICS

Abstract. In this article, we address the practical challenges encountered during the management and population of the institutional repository at the Faculty of Mathematics, University of Belgrade. Furthermore, we describe the development and implementation of an automated data extraction and conversion protocol designed to overcome low manual submission rates and ensure repository completeness.

Keywords: Institutional Repository, DSpace-CRIS, format transformation, BibTex, RIS, ORCID, eNauka.

An Institutional repository (IR) is digital database storing data and full text content of the institution's scientific production [1]. It includes articles from scientific journals, conference papers, books, book parts, textbooks, etc [2]. It can be built on various software platforms, with DSpace-CRIS being one of the most widely adopted options. DSpace-CRIS is comprehensive, free, and open-source Research Information Management System (CRIS/RIMS), developed and maintained by 4Science. CRIS/RIMS stands for Current Research Information Systems/Research Information Management Systems. It integrates data and full texts of an institution's scientific output, together with data on researchers, projects, and organizational units [3].

The institutional repository of the Faculty of Mathematics is accessible at <https://research.matf.bg.ac.rs/>. Developed using DSpace-CRIS version 5, it was created by Nikolina Vukša-Popović, who oversaw the initial data entry. The author assumed the role of administrator for the scientific collection in late September 2023, and at that moment, the repository contained about 1250 records. Despite establishing a workflow for data submission, the process heavily relied on individual researchers providing their records. This approach resulted in a low response rate, with only about 20 percent of the Faculty's staff submitting their data. To overcome this administrative bottleneck and ensure the repository's completeness, the author shifted focus from manual collection to an automated extraction protocol.

On the other hand, there is a similar repository on the state level called eNauka, which acquired data about scientific production from the various sources available previously. The platform is designed as a central database for tracking the scientific production of the entire Republic of Serbia, with institutional repositories harvest to the eNauka via OAI-PMH protocol [4]. Despite the previous announcements on establishing two-sided communication between eNauka and institutional repositories, which would synchronize eNauka and institutional repositories data, this functionality has not yet been implemented.

Data synchronization was further complicated by specific administrative dependencies. The workflow for establishing researcher profiles within the DSpace-CRIS platform strictly required staff members to submit their unique ORCID identifiers. Because a portion of the faculty staff failed to provide this information, their

institutional profiles could not be created. Consequently, during the data migration from eNauka, publications belonging to these unregistered authors could not be linked automatically, requiring additional manual verification and handling to manage the unassigned records.

To systematically approach the migration, we first evaluated the data compatibility across both systems. The eNauka platform offers several options for exporting scholarly metadata (Figure 1).

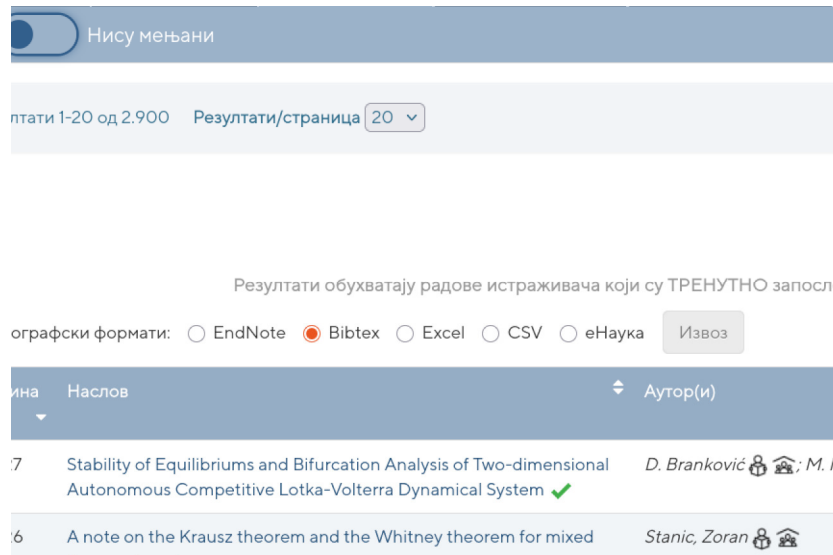


Figure 1: eNauka export format options

Concurrently, the DSpace-CRIS platform provides multiple native format specifications for importing bulk records into its database (Figure 2).

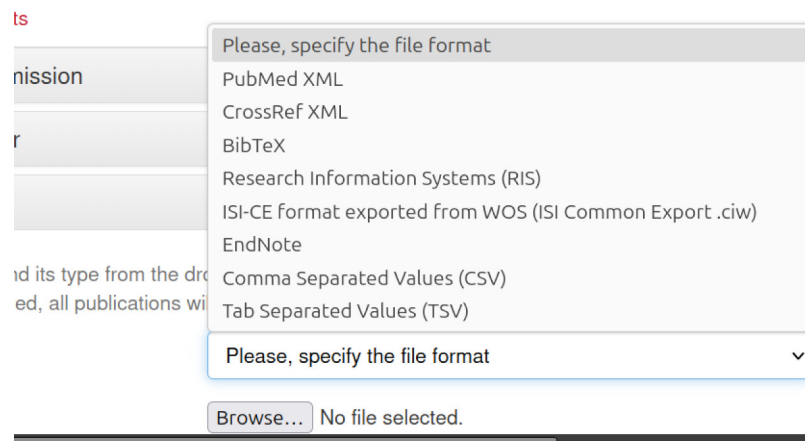


Figure 2: Import format options in DSpace-CRIS

The initial approach involved exporting records from eNauka in BibTeX format and subsequently importing them into the IR. However, this method revealed a significant technical limitation: the DSpace-CRIS import module failed to parse multiple authors into separate, repeatable metadata attributes, instead compressing all author names into a single field (Figure 3).

With the external lookup you can search both with the fullname and the ORCID iD

John W Montano and ... and Dragana J Ilic and ... and Andjelka

liation

Figure 3: Importing from Bibtex format

To address this issue, we tested the Research Information Systems (RIS) format for data import. This approach successfully resolved the parsing problem, correctly distributing individual authors into separate, repeatable fields (Figure 4).

The authors of this publication. With the external lookup you can search both with the fullname and the ORCID iD

Author	<input type="text" value="Montano, John W"/>
Affiliation	<input type="text"/>
Author	<input type="text" value="..."/>
Affiliation	<input type="text"/>
Author	<input type="text" value="Ilic, Dragana J"/>
Affiliation	<input type="text"/>
Author	<input type="text" value="Kovačević, Andjelka"/>
Affiliation	<input type="text"/>

Figure 4: Importing from RIS format

Consequently, a workflow was established: data is exported from eNauka in BibTeX format, converted to RIS using an established web-based converter available at <https://www.bibtex.com/c/bibtex-to-ris-converter/> (Figure 5), and then uploaded into the repository.

Your references in RIS

```
TY - MISC
AU - Montano, John W
AU - ...
AU - Ilic, Dragana J
AU - ...
AU - Kovacevic, Andjelka B
AU - ...
```

Figure 5: Converted record to RIS format

Another classification issue emerged during the migration of entries categorized as "conference papers." The imported metadata for these entries yielded formatting inconsistencies compared to other item types in the repository, preventing a uniform export. To achieve standardization, these entries were manually reclassified from

"Conference paper" to "Conference object." This matches the broader "conference object" typology utilized by eNauka since its inception. Ensuring this uniform typology within a single exported file is critical, as these records are heavily utilized for the generation of teaching staff promotion dossiers and official reporting to the relevant Ministries.

Conclusion

In this paper, we presented a practical and automated workflow developed to overcome the limitations of manual data curation in the institutional repository of the Faculty of Mathematics, University of Belgrade. Facing a low manual submission response rate of only 20% from the faculty staff, shifting the administration strategy toward a centralized data extraction protocol from the national eNauka platform proved to be a critical step in achieving repository completeness. By identifying and resolving specific metadata parsing issues—specifically through a standardized BibTeX-to-RIS conversion pipeline—we successfully migrated and accurately processed over 2,000 scholarly records.

Furthermore, the reclassification of inconsistent typologies, such as aligning "conference papers" into uniform "conference objects," has directly enhanced the repository's utility. The system now seamlessly supports the generation of clean, standardized datasets necessary for academic promotions, institutional evaluations, and official reporting to the relevant Ministries.

Future development of the repository will focus on two key areas. First, administrative efforts will continue toward collecting the remaining ORCID identifiers from the faculty staff, as providing these persistent identifiers remains a strict prerequisite for manual profile creation within our platform. Second, while the local infrastructure is already configured for standard OAI-PMH harvesting to the national eNauka platform, we aim to enhance our repository's ingestion capabilities. Our next technical step is to implement advanced integration modules that will allow the repository to automatically fetch and import scholarly metadata directly from global sources, specifically utilizing Crossref metadata harvesting and automated publication retrieval from researchers' verified ORCID profiles.

Finally, a parallel and continuous priority will be migrating and uploading the actual full-text documents for these records in accordance with publishers' copyright policies, transforming the platform from a metadata catalogue into a repository with the highest possible degree of open access. Ultimately, we hope that the insights and automated workflows presented in this article will serve as a practical reference for other institutions facing similar administrative and technical bottlenecks, thereby facilitating their own repository population and implementation processes.

References

- [1] Wikipedia, "Institutional repository on Wikipedia," [Online]. Available: https://en.wikipedia.org/wiki/Institutional_repository.
- [2] S. McGovern, N. Harnad, "Topic 4: Institutional repository success is dependent upon mandates," *Bulletin of the American Society for Information*, 2009, 35(4), 27–31
- [3] 4Science, "DSpace-CRIS on 4science," [Online]. Available: <https://4science.com/dspace-cris/>.

[4] O. Archive, "OAI-PMH protoco," [Online]. Available:
<https://www.openarchives.org/pmh/>.

nikola.scepancevic@matf.bg.ac.rs