

Enhancing Scholarly Knowledge Graph Interoperability through SKG-IF Integration in TeslaRIS

Dragan Ivanović¹, Ivan Mršulja², Miloš Popović³, and Lidija Ivanović⁴

¹ University of Novi Sad, Faculty of Technical Science, Novi Sad, Serbia
`dragan.ivanovic@uns.ac.rs`

² University of Novi Sad, Faculty of Technical Science, Novi Sad, Serbia
`ivan.mrsulja@uns.ac.rs`

³ University of Novi Sad, Faculty of Technical Science, Novi Sad, Serbia
`milospopovic@uns.ac.rs`

⁴ University of Novi Sad, Faculty of Education, Sombor, Serbia
`lidija.ivanovic@uns.ac.rs`

Abstract

Research information systems and scholarly knowledge graphs are central to Open Science, yet practical interoperability between them remains limited due to heterogeneous APIs and partially aligned semantic models. This paper presents the design, implementation, and validation of a full integration of the Scientific Knowledge Graph Interoperability Framework (SKG-IF) within TeslaRIS, an open-source research information system deployed at the University of Novi Sad. We describe data model alignment using SSSOM-based crosswalks, a production SKG-IF 1.1.0 server exposing more than 125,000 entities, and a configurable client for harvesting external SKG-IF sources. The work demonstrates how institutional CRIS platforms can participate in a federated scholarly knowledge graph ecosystem while supporting FAIR principles and alignment with the European Open Science Cloud.

Keywords: Scholarly Knowledge Graphs, Interoperability, CRIS, SKG-IF, Open Science, FAIR, TeslaRIS, EOSC.

1 Introduction

Research information systems (CRIS) and scholarly knowledge graphs (SKGs) play a central role in the management, discovery, and dissemination of research outputs. Over the past decade, numerous national, European, and global initiatives (such as institutional CRIS platforms, OpenAIRE Graph, OpenCitations, and domain-specific knowledge graphs) have emerged to support Open Science and evidence-based research assessment. While these initiatives increasingly rely on graph-based representations of scholarly information, their practical interoperability remains limited. Data is often exposed through heterogeneous APIs, proprietary schemas, or partially aligned semantic models, resulting in fragmentation and duplication of effort.

The lack of interoperability between SKGs directly affects the efficiency and sustainability of research infrastructures. Institutions are required to maintain multiple export formats, aggregators must implement custom harvesting pipelines, and the reuse of structured scholarly metadata across systems is hindered. These challenges are particularly relevant in the CRIS community, where institutional systems are expected to seamlessly exchange data with national and European-level infrastructures while preserving semantic richness and provenance.

The Scientific Knowledge Graph Interoperability Framework (SKG-IF) has recently been proposed as a standardized, API-driven approach for exposing and consuming scholarly knowledge graph data. SKG-IF defines a common data model, serialization formats, and filtering mechanisms aimed at enabling machine-to-machine interoperability between heterogeneous scholarly platforms. However, at the time of this work, practical, production-level implementations of SKG-IF within operational CRIS platforms were scarce.

This paper presents the design, implementation, and validation of a full SKG-IF integration within TeslaRIS, an open-source research information system developed and deployed at the University of Novi Sad (Mršulja et al., 2024). The work demonstrates how SKG-IF can be operationalized in a real institutional context, highlights the technical and semantic challenges encountered, and discusses the implications for CRIS interoperability within the European Open Science ecosystem.

2 Related Work

Scholarly knowledge graphs model research entities (publications, datasets, persons, organizations, venues, and projects) and their relationships using graph-based structures. Prominent examples include OpenAIRE Graph, Open Research Knowledge Graph, and OpenCitations. In parallel, CRIS platforms such as Pure, Converis, and various open-source solutions provide institutional-level research information management, often aligned with standards such as CERIF.

Despite overlapping scopes, SKGs and CRIS platforms have evolved with different design priorities. CRIS systems emphasize institutional workflows, validation, and reporting, while global SKGs prioritize large-scale aggregation and analytics. Existing interoperability approaches typically rely on batch exports (e.g., CERIF XML, CSV) or bespoke REST APIs, which limits real-time exchange and semantic alignment.

SKG-IF addresses these limitations by proposing a common conceptual model for core scholarly entities and relationships including standardized API endpoints for querying and filtering entities. Moreover, SKG-IF should support semantic web technologies representations, including JSON-LD and RDF serializations. At the end, there are clear rules for pagination, content negotiation, and extensibility.

The data model uses Semantic Web standards and reuses existing ontologies, including the Semantic Publishing and Referencing (SPAR) Ontologies (Peroni and Shotton, 2018) and the Provenance Ontology (PROV-O) (Lebo et al., 2013). Aligning CRIS platforms with SKG-IF allows institutional systems to function as integral components of a distributed network of interoperable scholarly knowledge graphs. As discussed by Aryani et al. (2020), improved interoperability helps reduce duplication of effort and supports a better use of complementarities across different SKGs.

3 TeslaRIS Platform

The development of TeslaRIS represents the culmination of more than fifteen years of research and development efforts in the field of research information management at the University of Novi Sad (Surla et al., 2013). The initiative originated in 2009 through the DOSIRD UNS project, which was launched to address the lack of comprehensive software support for research activities within the university. While information systems supporting educational and administrative processes were already well established, the research domain required a dedicated infrastructure capable of managing scholarly outputs, institutional reporting, and research evaluation processes.

One of the key outcomes of the DOSIRD UNS project was the development of the CRIS UNS platform, which became operational in 2010. The system was based on a CERIF-compatible data model enriched with MARC 21 bibliographic records (Ivanović et al., 2011, 2010), enabling detailed descriptions of research entities and supporting interoperability with standards such as Dublin Core and ETD-MS (Ivanović et al., 2012). Over time, the platform evolved into a central repository of research information, incorporating a digital library of doctoral dissertations (Ivanović et al., 2013) and extensions required for the evaluation of scientific results according to Serbian national regulations (Dimić Surla et al., 2012; Nikolić et al., 2015). During more than a decade of operation, the system accumulated over 40,000 research-related entities and continuously expanded its functionality.

Despite its success, the long-term evolution of CRIS UNS introduced significant technical debt and architectural limitations. Advances in software engineering technologies, the emergence of new web development frameworks, and changes in scholarly communication practices highlighted the need for a comprehensive redesign. As a result, the TeslaRIS project was initiated as a reengineering effort aimed at creating a modern, scalable, and reusable research information system. Unlike its predecessor, TeslaRIS was conceived as an open-source platform that could be adopted by institutions beyond the University of Novi Sad, with CRIS UNS becoming only one possible customization of the broader framework.

The TeslaRIS data model follows the fundamental principles of the refactored CERIF model while introducing additional flexibility and improvements. Research outputs are connected to contributors through explicit contribution entities that preserve historical affiliation information. This approach enables accurate attribution of publications to institutions even when researchers change affiliations over time. Furthermore, multilingual content is managed through dedicated entities that support seamless representation of metadata in multiple languages, thereby facilitating internationalization and broader interoperability.

From a technological perspective, TeslaRIS is built as a modular monolithic application using the Spring Boot framework (Figure 1). The architecture combines modularity with the simplicity of deployment and maintenance characteristic of monolithic systems. PostgreSQL serves as the primary relational database for storing structured research information, while Elasticsearch provides advanced indexing and search capabilities. Document files are stored using MinIO object storage, ensuring scalability and reliability for digital assets.

The platform has been used at University of Novi Sad (<https://cris.uns.ac.rs/en>). The UI of TeslaRIS is modern and includes numerous representations of research information (Figure 2).

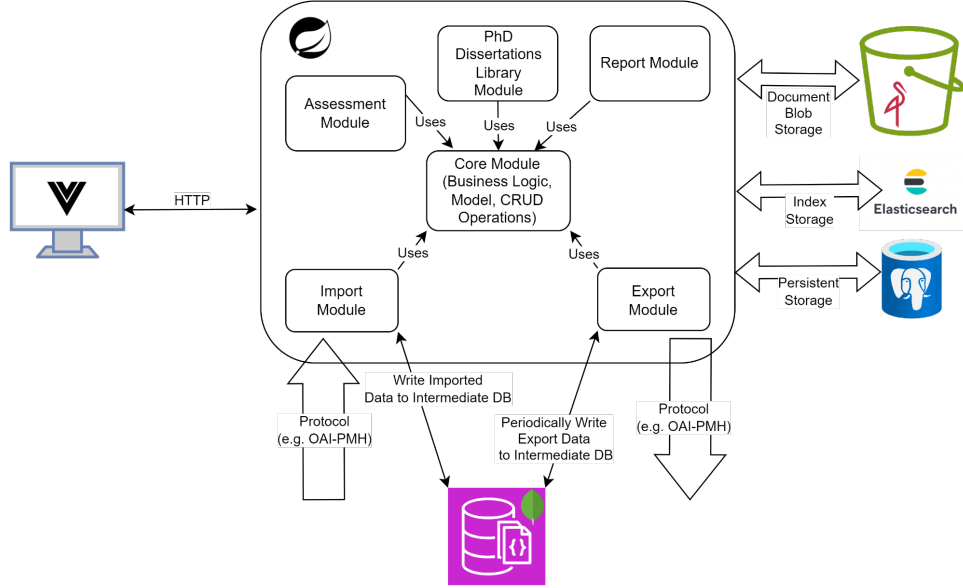


Figure 1: TeslaRIS architecture.

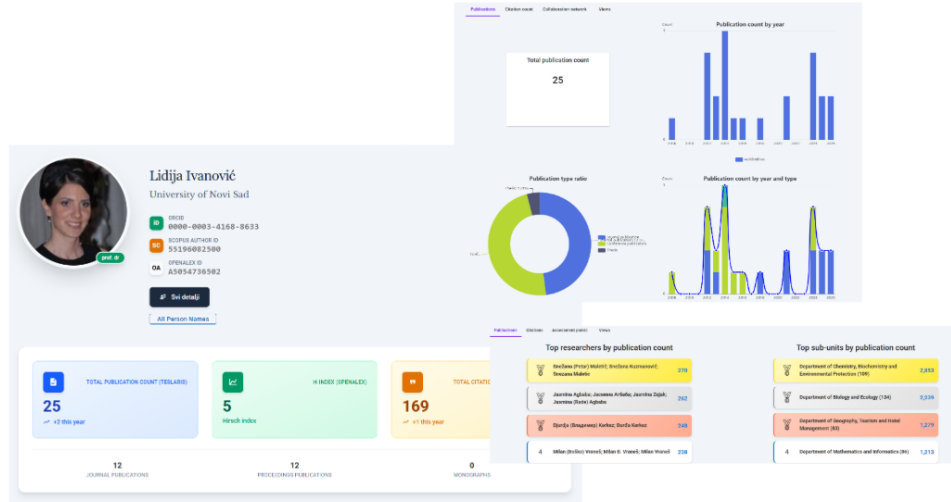


Figure 2: TeslaRIS in action.

4 Data Model Alignment

A prerequisite for SKG-IF implementation was the alignment of the TeslaRIS internal data model with the SKG-IF data model. TeslaRIS already maintains a rich object model covering research products, contributors, affiliations, and venues. However, semantic gaps and structural differences required careful analysis.

The alignment process involved:

- Mapping TeslaRIS entities to corresponding SKG-IF entity types
- Harmonizing identifiers, controlled vocabularies, and relationship semantics
- Extending the internal model where necessary to ensure lossless transformation
- Implementing the required API discovery and healthcheck infrastructure, proposed by API specification

The resulting model alignment was documented in UML diagrams and formalized through a crosswalk expressed in the Simple Standard for Sharing Ontological Mappings (SSSOM). This crosswalk serves as a machine-readable and human-readable specification of semantic equivalence between the two models.

The SSSOM-based crosswalk covers key entities, including Agent, Person, Organization Unit, Venue, Journal, Conference Proceedings, and Research Product subtypes (e.g., journal articles and conference papers). A dedicated GitHub Wiki page has been created to publish and document the mappings: <https://github.com/sci2zero/TeslaRIS-backend/wiki/Crosswalks>. By explicitly documenting these mappings, the implementation ensures transparency, maintainability, and extensibility. Figure 3 presents part of the Person entity mapping.

This approach also supports future evolution of either model, as changes can be localized within the mapping layer rather than requiring ad hoc code modifications.

5 SKG-IF Server Implementation

The SKG-IF API was implemented as a native component of the TeslaRIS backend. The implementation fully complies with SKG-IF specification version 1.1.0 and includes:

- Standard SKG-IF endpoints for entity retrieval
- Support for pagination and filtering mechanisms defined by the framework
- Content negotiation with support for JSON-LD (default) and RDF/Turtle serializations
- Proper handling of identifiers and links between entities
- Follows SKG-IF context and vocabulary standards
- Additional TeslaRIS extensions, including date range filtering and output format selection (for single entity export) are fully backward compatible

To minimize response latency and avoid repeated execution of complex relational queries, the SKG-IF API reuses the intermediate MongoDB document representation that was originally developed for TeslaRIS OAI-PMH exports Mršulja et al. (2024). During the indexing process, research entities are transformed into denormalized JSON documents containing all metadata required for SKG-IF serialization. API requests are therefore served directly from MongoDB, eliminating the need for expensive joins over the PostgreSQL transactional database. This architecture provides low-latency responses while keeping the API isolated from the operational workload of the primary data store. Additionally, this avoids duplication of transformation logic and allows both interoperability services to share a common denormalized representation of scholarly metadata.

Agent

foaf:Person → teslaris:Person

- **Match:** `skos:exactMatch`
- **Justification:** `semapv:ManualMappingCuration`
- **Condition:** In JSON there should be `entity_type="person"`

skgif:local_identifier → teslaris:localIdentifier

- **Match:** `skos:exactMatch`
- **Justification:** `semapv:LexicalMatching`
- **Comment:** Only applicable if format is JSON; in RDF it's part of the URI

datacite:hasIdentifier[scheme=orcid] → teslaris:orcid

- **Match:** `skos:narrowMatch`
- **Justification:** `semapv:ManualMappingCuration`
- **Condition:** `identifier.scheme="orcid"`
- **Comment:** `identifier.value (literal:hasLiteralValue)` maps only when scheme is `orcid` (`datacite:usesIdentifierScheme datacite:orcid`)

foaf:givenName → teslaris:firstname

- **Match:** `skos:exactMatch`
- **Justification:** `semapv:LexicalMatching`

Figure 3: Example of mappings (Person).

The API exposes more than 125,000 scholarly entities stored in the TeslaRIS instance at the University of Novi Sad, making them accessible to external systems in a standardized, machine-actionable manner. The endpoint is available at <https://cris.uns.ac.rs/api/skg-if/>. As of June 8, 2026, there are:

- 39,008 research products
- 82,444 venues
- 12,762 persons
- 269 organisations

One example of filtering research product is by contributor, for instance to fetch all products published by the researcher Lidija Ivanović (<https://cris.uns.ac.rs/sr/persons/3066>), the following filter can be used: https://cris.uns.ac.rs/api/skg-if/product?page=0&page_size=50&filter=contributions.by:3066, or if we want to use a global identifier such as ORCID (0000-0003-4168-8633), Scopus Author ID (55196082500), OpenAlex ID (A5054736502), we can use the following request filters:

- https://cris.uns.ac.rs/api/skg-if/product?page=0&page_size=50&filter=contributions.by.identifiers.value:0000-0003-4168-8633
- https://cris.uns.ac.rs/api/skg-if/product?page=0&page_size=50&filter=contributions.by.identifiers.value:55196082500
- https://cris.uns.ac.rs/api/skg-if/product?page=0&page_size=50&filter=contributions.by.identifiers.value:A5054736502

Moreover, we can list all articles published in a journal (*Scientometrics*) by using the following request filter:

- https://cris.uns.ac.rs/api/skg-if/product?page=0&page_size=50&filter=biblio.in:Scientometrics.

In addition to individual filtering criteria, the TeslaRIS implementation supports the composition of multiple filter expressions within a single request. Multiple filters are combined using a comma-separated syntax in the filter parameter, allowing clients to progressively narrow the result set without issuing multiple API requests.

This functionality enables more expressive scholarly queries by combining constraints on different metadata dimensions. For example, it is possible to retrieve only publications authored by a specific researcher that were published in a particular journal:

- https://cris.uns.ac.rs/api/skg-if/product?page=0&page_size=50&filter=biblio.in:Scientometrics,contributions.by:3066

In this example, the first filter limits results to publications appearing in the journal *Scientometrics*, while the second restricts the result set to publications authored by the researcher whose internal TeslaRIS identifier is 3066. Only products satisfying both criteria are returned.

The implementation is designed to support arbitrary combinations of compatible filters, allowing clients to formulate increasingly precise queries while remaining fully compliant with the SKG-IF filtering model. Each filter expression is translated into an individual database predicate, after which all predicates are combined using logical conjunction (AND) before query execution.

For a certain research output we can request response in JSON or Turtle format, for instance:

- <https://cris.uns.ac.rs/api/skg-if/product/32467?format=json>
- <https://cris.uns.ac.rs/api/skg-if/product/32467?format=ttl>

6 SKG-IF Client Implementation

In addition to exposing data, TeslaRIS was extended to consume data from external SKG-IF endpoints. A configurable harvesting mechanism allows administrators to define external sources, specify filtering parameters, and control identifier management to avoid conflicts, all using simple JSON-based configuration files.

Listing 1 shows an example source definition in the system configuration file.

After defining sources in respective configuration files, they will seamlessly appear in the harvester UI, ready for use (Figure 4).

Once set-up, the scheduled task will appear in the “Scheduled Harvests” tab. The result of this harvest will be temporarily stored in a MongoDB intermediate database. Every author

Listing 1: Configuration of a source.

```

{
  "sources": [
    {
      "sourceName": "CRIS UNS",
      "baseUrl": "https://cris.uns.ac.rs/api/skg-if",
      "metadataFormatParameter": "format",
      "dateFromFilterParam": "dateFrom",
      "dateToFilterParam": "dateTo",
      "additionalFilters": "product_type:literature",
      "converterClass": "SKGIFConverter",
      "sourceIdentifierPrefix": "TESLARIS"
    }
  ]
}

```

Figure 4: UI for scheduling harvesting.

from the publication that has an user profile on our platform will receive a notification stating that they can perform the loading process.

After the researcher clicks *Load documents*, the loading process guides the researcher through a wizard that performs semi-automated loading of data. First, the user has to bind every author from the imported publication to an author in our system. The system performs this step automatically using the import platform's unique identifier (e.g. Scopus Author ID, ORCID ID etc.). If it fails to do so, our database is queried by author name and the user is given a list of potential candidates to manually choose from. Should the user not find an adequate match or should our database query return no results, the new researcher is automatically created and the process is continued to the next author until we reach the end of the list. The important detail here is that once this step is completed with a certain author, the platform's unique ID will be automatically bound to said researcher, ensuring that next time this process is completed automatically by the system, without additional user interaction. After that, the same process

is repeated for authors' affiliations and journal/proceedings that the publication is published in. In the end, the user is presented with other publication metadata, such as description, keywords etc., so they can edit all other details about the publication. Submitting this last form means that publication will be successfully loaded into our system and the user can continue to the next one.

This dual role (provider and consumer) positions TeslaRIS as an active participant in a distributed SKG ecosystem rather than a passive data source.

7 Results & Discussion

The SKG-IF integration was deployed in the production TeslaRIS instance of the University of Novi Sad. At the time of reporting (June 8th, 2026), the exposed dataset included more than one hundred thousand entities (research products, venues, persons, and organizations) combined.

Full SKG-IF 1.1.0 compliance is provided through the `/skg-if` discovery endpoint, which offers automatic API documentation and service discovery. The endpoint includes:

- **API Discovery** (GET `/skg-if`): Returns complete SKG-IF metadata including API version (1.1.0), server information, supported collections (Products, Persons, Organizations, Venues), filtering capabilities, pagination options, and licensing details. The discovery response dynamically reports entity counts from the database and lists all supported filtering parameters for each collection.
- **Health Check** (GET `/skg-if/echo`): Provides real-time system status monitoring including MongoDB connectivity, API version information, and Git commit details. For security reasons, detailed metrics are not exposed through this endpoint.
- **Comprehensive Collections**: Four main resource collections are exposed: Products (research outputs including publications, datasets, and products with over 100K entities), Persons (researchers with their affiliations), Organizations (university organizational hierarchy), and Venues (publishing gateways such as journals, proceedings, and monographs).
- **Advanced Filtering**: Each collection supports sophisticated filtering capabilities including temporal filtering, contributor-based queries, and attribute-based selection.
- **Production Features**: The implementation includes proper context resolution, license information (GPL-3.0), contact details, and extension metadata about supported harvesting sources and FAIR compliance.

The system automatically reports its capabilities through the discovery endpoint, enabling client applications to dynamically adapt to available features and filter options without hard-coded configuration.

To validate interoperability, two TeslaRIS instances were configured in a provider-consumer setup using SKG-IF. This experiment demonstrated successful round-trip data exchange, confirming that both export and import functionalities operate correctly in practice.

The decision to serve SKG-IF responses from the denormalized MongoDB representation resulted in consistently low response times, since API requests require only document retrieval and serialization rather than executing complex joins over the relational data model.

The full implementation, including source code, documentation, and mappings, has been released as open source. This ensures reproducibility and enables other institutions to adopt or adapt the solution with minimal effort.

The implementation revealed several important insights. First, while the SKG-IF specification provides a solid foundation, aligning existing CRIS data models requires significant semantic analysis and engineering effort. Second, the absence of widely adopted SKG-IF endpoints among major aggregators currently limits immediate network effects. Nevertheless, the presented implementation lowers the barrier for future adoption by providing a concrete reference architecture.

From a CRIS perspective, SKG-IF offers a promising complement to established standards such as CERIF. Rather than replacing existing reporting-oriented formats, SKG-IF enables dynamic, API-based interoperability suitable for modern, graph-driven research infrastructures.

This work directly addresses key themes of the euroCRIS community, including interoperability, Open Science, and integration with the European Open Science Cloud (EOSC). By operationalizing SKG-IF within an institutional CRIS, the project demonstrates how local research information systems can align with the EOSC Interoperability Framework and participate in a federated European knowledge space.

The open-source nature of TeslaRIS and the documented SKG-IF integration provide a reusable blueprint for other institutions, particularly those seeking vendor-neutral solutions. The approach supports FAIR principles at the metadata level, with particular emphasis on interoperability and reusability.

8 Conclusions

This paper presented a comprehensive, production-level implementation of the SKG Interoperability Framework within the TeslaRIS research information system. The results demonstrate the technical feasibility and practical benefits of SKG-IF for enhancing CRIS interoperability and integrating institutional data into a broader scholarly knowledge graph ecosystem.

Future work will focus on promoting adoption within the CRIS and SKG communities, integrating with external SKG-IF-enabled platforms as they emerge, and refining semantic mappings based on community feedback. Broader uptake of SKG-IF has the potential to significantly reduce fragmentation in scholarly infrastructures and strengthen the foundations of Open Science in Europe and beyond.

An additional direction of future work is the development of interactive graph-based exploration of scholarly entities. The graph-oriented data model exposed through SKG-IF naturally supports incremental retrieval of interconnected entities. A client application can first request a single entity together with its immediate relationships and then progressively retrieve additional neighboring entities as the user expands different parts of the graph. This on-demand traversal minimizes unnecessary data transfer while enabling responsive visualization of large scholarly knowledge graphs.

TeslaRIS already includes a graph visualization interface for exploring researcher affiliations and collaboration networks. Building upon the SKG-IF integration, this visualization can be extended to navigate the complete scholarly ecosystem, including research products, publication venues, organizations, contributors, grants, and research topics exposed through the standardized API. Such an approach would allow users to seamlessly explore scholarly knowledge across institutional boundaries while relying on a common interoperability framework rather than platform-specific interfaces.

Beyond visualization, this capability opens opportunities for richer semantic exploration, interactive discovery of collaboration patterns, and cross-institutional navigation through interconnected scholarly knowledge graphs, further demonstrating the value of SKG-IF as both an interoperability protocol and a foundation for next-generation research information services.

References

- Aryani, A., Fenner, M., Manghi, P., Mannocci, A., and Stocker, M. (2020). Open science graphs must interoperate! In *Scientific Knowledge Graphs at the International Conference on Theory and Practice of Digital Libraries 2020 (TPDL) (SKG)*.
- Dimić Surla, B., Segedinac, M., and Ivanović, D. (2012). A BIBO ontology extension for evaluation of scientific research results. In *Proceedings of the Fifth Balkan Conference in Informatics*, pages 275–278.
- Ivanović, D., Milosavljević, G., Milosavljević, B., and Surla, D. (2010). A CERIF-compatible research management system based on the MARC 21 format. *Program: electronic library and information systems*, 44(3):229–251.
- Ivanović, D., Surla, D., and Konjović, Z. (2011). CERIF compatible data model based on MARC 21 format. *The Electronic Library*, 29(1):52–70.
- Ivanović, L., Ivanović, D., and Surla, D. (2012). A data model of theses and dissertations compatible with CERIF, Dublin Core and EDT-MS. *Online Information Review*, 36(4):548–567.
- Ivanović, L., Ivanović, D., Surla, D., and Konjović, Z. (2013). User interface of web application for searching PhD dissertations of the university of novi sad. In *2013 IEEE 11th International Symposium on Intelligent Systems and Informatics (SISY)*, pages 117–122. IEEE.
- Lebo, T., Sahoo, S., McGuinness, D. L., Belhajjame, K., Cheney, J., Corsar, D., Garijo, D., Soiland-Reyes, S., Zednik, S., and Zhao, J. (2013). PROV-O: The PROV ontology. W3c recommendation, World Wide Web Consortium.
- Mršulja, I., Popović, M., Ivanović, D., Ivanović, L., and Surla, D. (2024). Reengineering of cris at university of novi sad. In *Procedia Computer Science (CRIS 2024)*, volume 249, pages 85–93. Elsevier.
- Nikolić, S., Konjović, Z., Penca, V., Ivanović, D., and Surla, D. (2015). A CERIF compatible CRIS-UNS model extension for assessment of conference papers. *Acta Polytechnica Hungarica*, 12(7):129–148.
- Peroni, S. and Shotton, D. (2018). The SPAR ontologies. In *The Semantic Web – ISWC 2018*, volume 11137 of *Lecture Notes in Computer Science*, pages 119–136. Springer, Cham.
- Surla, D., Ivanović, D., and Konjović, Z. (2013). Development of the software system CRIS UNS. In *2013 IEEE 11th International Symposium on Intelligent Systems and Informatics (SISY)*, pages 111–116. IEEE.