DCPAPERS

DCMI 2026

Proceedings of the International Conference on Dublin Core and Metadata Applications 2026

ISSN: 1939-1366

Beyond Mapping: A Semantic Transformation Approach from KORMARC and MODS to BIBFRAME

Seungmin Lee [1]

[1] Chung-Ang University

Abstract

MARC- and MODS-based bibliographic data, due to their record-based structure, have limitations in supporting entity identification and semantic relationships in linked data environments. Existing approaches to converting these formats into BIBFRAME primarily rely on field-level mappings, which fail to capture bibliographic meaning and relationships adequately. To address this issue, this study proposes a semantic-based transformation framework that reinterprets bibliographic data at the level of semantic units. The framework includes a normalization model and a transformation module consisting of semantic analysis, entity extraction, relationship generation, and RDF transformation. This approach enables the reconstruction of bibliographic data into an entity–relationship structure, facilitating their conversion into BIBFRAME while preserving semantic integrity and enhancing interoperability and reusability.

1. Introduction

The bibliographic environment is rapidly shifting from traditional record-based structures to linked data models centered on entities and relationships. [1] This transition represents a paradigm shift in which bibliographic data are no longer treated as isolated records, but as interconnected and reusable knowledge resources in the web environment. In particular, BIBFRAME has emerged as a key framework for restructuring MARC-based data into a model suitable for the Semantic Web. [1]

In response, libraries have undertaken various efforts to convert existing MARC and MODS data into BIBFRAME. However, these attempts have not fully achieved their intended goals and have shown limitations in practical adoption. [2] This is largely due to the reliance on field-level mappings that focus on syntactic conversion rather than semantic reconstruction. [3]

The fundamental limitation of these approaches lies in the structural differences between record-based and entity–relationship-based models. MARC and MODS integrate bibliographic elements within a single record, whereas BIBFRAME distributes information across entities and expresses relationships explicitly. When these differences are not adequately addressed, issues such as semantic loss, incomplete relationships, and duplication of entities arise. [4]

To address these challenges, this study proposes a semantic-based transformation framework that reinterprets bibliographic data as units of meaning and reconstructs them into an entity–relationship structure. Through this approach, the study aims to

Full Paper

Available Aug 01, 2026

Correspondence to
Seungmin Lee
ableman@cau.ac.kr

enable the conversion of KORMARC and MODS data into BIBFRAME while preserving semantic integrity and enhancing interoperability.

2. Theoretical background

2.1. Limitations of record-based bibliographic data structures

Traditional bibliographic data have been structured primarily in record-based formats such as MARC, where multiple descriptive elements are integrated within a single record. [5] While this structure has been effective in conventional cataloging environments, much of the meaning remains implicit and dependent on human interpretation rather than being explicitly machine-readable.

Such record-based structures have limitations in representing relationships among bibliographic elements. Because relationships are often embedded implicitly within records, it is difficult for machines to interpret and utilize them in linked data environments that require explicit, relationship-oriented models.

In addition, the record-centric approach restricts entity identification and interoperability. The same entity may be repeatedly described across multiple records without a unified identifier, and differences in metadata formats such as MARC and MODS can lead to semantic inconsistencies during transformation.

Overall, these limitations highlight the need to move beyond record-based structures and to reconstruct bibliographic data as entity- and relationship-based models that support semantic interpretation and interoperability. [6]

2.2. BIBFRAME and linked data

BIBFRAME, developed by the Library of Congress in 2012, is a key framework designed to transform traditional MARC-based bibliographic data into a format suitable for linked data environments. It is based on an entity-oriented data model, representing a fundamental departure from conventional systems that describe bibliographic information within a single record. [7]

The core structure of BIBFRAME consists of primary entities such as Work, Instance, and Item. Work represents the conceptual level of an intellectual or artistic creation, Instance refers to a specific embodiment of that work, and Item denotes an individual physical or digital copy. Unlike traditional record-based approaches, this structure enables the hierarchical representation of complex bibliographic relationships, such as multiple expressions and manifestations of a single work.

In addition, BIBFRAME adopts an RDF-based data model, in which all bibliographic elements are expressed as triples. In this process, each entity is assigned a unique URI for identification, and relationships among entities are explicitly defined through properties. [8] This structure, unlike string-based record representations, allows machines to directly interpret connections among data and facilitates integration with external datasets.

Importantly, BIBFRAME goes beyond a mere change in data structure and aims at the semantic reconfiguration of bibliographic data. That is, it decomposes the implicit meanings embedded in MARC fields into entities and relationships and re-expresses them explicitly, thereby enhancing the precision of semantic representation. [9] This approach provides a foundation for integrating bibliographic data not as isolated records, but as part of a knowledge graph.

In summary, BIBFRAME serves as a core model for reconstructing bibliographic data in linked data environments, offering significant potential to improve data identification, connectivity, and scalability. However, to fully realize this potential, it is necessary to move beyond simple technical transformation and adopt semantic-based transformation strategies that reflect the underlying meaning structure of bibliographic data.

2.3. Limitations of existing transformation approaches

Research on transforming existing bibliographic data, such as MARC and MODS, into BIBFRAME has been conducted from various perspectives. However, many of these studies rely heavily on syntactic mapping based on field-to-field correspondence, which fails to adequately reflect the underlying semantic structure of bibliographic data. [10] Even when the same field is used, its meaning may vary depending on structural context. Therefore simple mapping approaches that do not account for such contextual differences are likely to result in semantic distortion or information loss during the transformation process.

Another limitation of existing approaches is the insufficient transition to an entity-based structure. Although BIBFRAME is designed as a model centered on entities and relationships, many transformation methods remain at the level of converting existing record structures into RDF format without fundamentally restructuring the data. As a result, hierarchical relationships among entities such as Work, Instance, and Item are not clearly represented, and the structural advantages of BIBFRAME are not fully realized.

The lack of semantic relationship modeling is also a major concern. While MARC and MODS often represent relationships implicitly, BIBFRAME requires explicit definition of relationships. However, existing transformation studies have not sufficiently analyzed or defined the types and directions of relationships among bibliographic elements such as authors, subjects, and contributors. Consequently, the resulting data may conform to RDF syntax but lack meaningful semantic connectivity in practice.

These limitations suggest that a purely syntactic transformation approach is insufficient to fully realize the potential of BIBFRAME-based bibliographic data. Therefore, it is necessary to move beyond field-based mapping and adopt a semantic-based transformation approach that reconstructs entities and relationships in accordance with the underlying meaning structure of bibliographic data.

3. Design of a semantic-based BIBFRAME transformation framework

3.1. Necessity of a semantic-based transformation module

This study proposes a semantic-based transformation framework that goes beyond conventional field-level mapping approaches in order to convert MARC- and MODS-based bibliographic data into the BIBFRAME structure.

Transforming bibliographic records from KORMARC, the Korean MARC standard, and MODS into BIBFRAME requires more than simple field-to-field correspondence; it demands an approach that reflects the underlying semantic structure of bibliographic data. Existing transformation methods typically perform field-level mappings while preserving the original record structure, which limits their ability to adequately represent relationships among entities and the semantic context embedded in the data. In particular, while MARC and MODS integrate diverse bibliographic information within a single record, BIBFRAME distributes data across entities and expresses them through explicit relationships. This fundamental structural difference necessitates not merely a format conversion, but a comprehensive reconstruction of bibliographic data.

Accordingly, this study adopts bibliographic normalization as a prerequisite for designing a transformation module from KORMARC and MODS to BIBFRAME. Bibliographic data are reinterpreted as semantic units and systematically reorganized through a normalization process. The proposed framework positions bibliographic normalization as a core preliminary stage prior to transformation, functioning as a middleware layer that resolves structural differences among heterogeneous bibliographic record formats and reorganizes them into a unified, meaning-based structure.

3.2. Concept of bibliographic normalization

Bibliographic normalization is the process of eliminating structural differences among heterogeneous metadata formats and reorganizing them into a unified, meaning-based structure. It functions as an intermediate layer (middleware) performed prior to the design of the transformation module, providing a foundation for converting data into an entity-based structure while preserving their semantic integrity.

In traditional field- and tag-based approaches, elements with the same meaning may be represented differently depending on the metadata format, while identical fields may convey different meanings depending on context. To address these issues, this study redefines bibliographic elements not at the level of fields, but as units of bibliographic meaning, and reconstructs the data structure accordingly.

This normalization process goes beyond simple data organization, playing a critical role in establishing semantic-based transformation criteria that can be aligned with the Class–Property–Relationship structure of BIBFRAME.

3.3. Five-layer bibliographic normalization model

To design a semantic-based transformation module, this study proposes a bibliographic normalization model consisting of the following five layers.

(1) Record Structure Layer

The record structure layer does not directly replicate the physical structure of KORMARC and MODS, but instead focuses on extracting the bibliographic description units embedded within those structures. By moving away from traditional field- and tag-based approaches, this layer decomposes record elements into semantic units, providing a foundation for subsequent reconstruction in higher layers.

(2) Bibliographic Entity Layer

At this layer, bibliographic data are reorganized into entities corresponding to the core structure of BIBFRAME, namely Work, Instance, and Item. A single record is decomposed into multiple entities, and hierarchical relationships among these entities are established. This process clarifies how bibliographic information is connected at different levels and represents a key step in transitioning to an entity-based data model.

(3) Semantic Function Layer

The semantic function layer classifies bibliographic elements based on their functional roles rather than their original fields. For example, elements are categorized into functions such as identifier, title, agent, subject, description, and relation. This function-based classification enables the integration of heterogeneous structures from KORMARC and MODS and provides a foundation for mapping these elements to BIBFRAME RDF properties.

(4) Value Representation Layer

The value representation layer addresses the issue that a single bibliographic element may have multiple forms of expression, such as strings, authorized access points, or identifiers. In this study, values are systematically categorized as literal, coded value, identifier, URI-linked entity, and note. This classification clarifies the meaning of values and minimizes information loss during the BIBFRAME transformation process.

(5) Relationship Network Layer

The relationship network layer reconstructs relationships among bibliographic elements in accordance with the property and relation structures of BIBFRAME. Relationships that were implicitly represented in traditional records are explicitly defined, and connections among entities are expanded into a network structure. This layer is designed to support not only relationships within a single record but also connections with external bibliographic data.

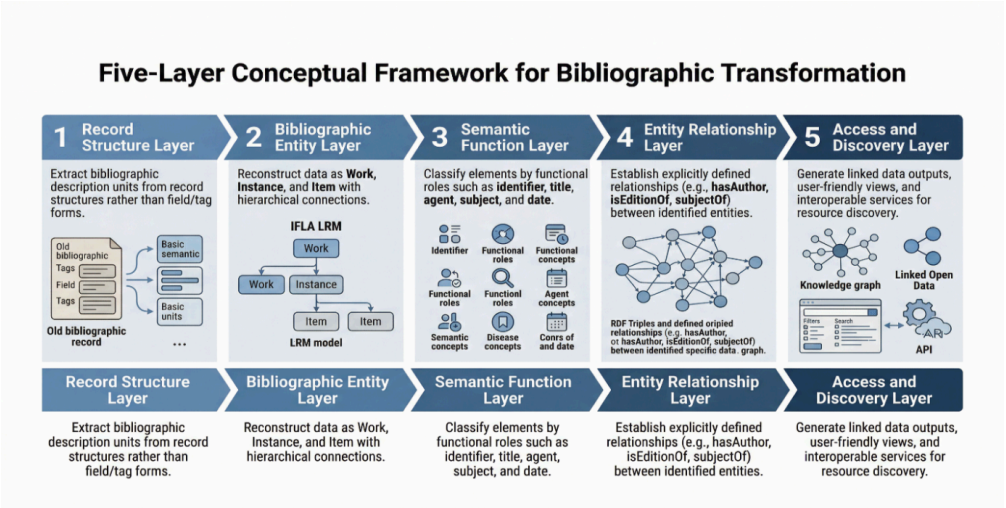

Figure 1. *A Five-Layer Bibliographic Normalization Model*

3.4. Role and significance of the normalization model

The five-layer bibliographic normalization model proposed in this study provides a fundamental framework for integrating the heterogeneous structures of KORMARC and MODS and transforming them into the entity–relationship structure of BIBFRAME. In particular, this model defines the transformation target not at the level of fields or tags, but as units of bibliographic meaning, thereby addressing the problem of semantic loss that often occurs in conventional mechanical mapping approaches.

Furthermore, by simultaneously reflecting both entity-based and relationship-oriented structures, the model establishes a foundation for extending bibliographic data into forms suitable for use in linked data environments.

Ultimately, the proposed normalization model goes beyond simple data transformation and serves as a starting point for the structural transition toward reconstructing bibliographic data as a knowledge graph.

4. Design of a semantic-based transformation module structure

4.1. Basic structure of the transformation module

This study designs a semantic-based transformation module for converting KORMARC- and MODS-based bibliographic data into the BIBFRAME structure, based on the bibliographic normalization model proposed earlier. Conventional transformation approaches have relied on syntactic mapping based on field-to-field correspondence, which has been insufficient for capturing relationships among entities and the underlying semantic meaning of bibliographic data. Accordingly, this study defines the transformation module not merely as a conversion tool, but as a structural system for interpreting and reconstructing bibliographic meaning.

The proposed transformation module is based on three key design principles. First, it adopts an entity-based structure by decomposing bibliographic data into entities such as Work, Instance, and Item, and assigning URI-based identifiers to ensure data reusability

and connectivity. Second, it applies a relationship-oriented model by explicitly defining relationships among various bibliographic elements, including authors, subjects, and publication information. Third, it employs semantic mapping, in which transformation is performed based on units of meaning rather than fields, aligning them with the BIBFRAME structure.

The proposed module operates as a modular processing system that takes normalized bibliographic structures as input and transforms them into BIBFRAME data in RDF format through semantic analysis and structural reconstruction. The overall transformation process follows a workflow of input–analysis–transformation–output, where each stage operates independently while remaining structurally interconnected. In particular, the module is designed to process heterogeneous metadata formats such as KORMARC and MODS in an integrated manner. Furthermore, instead of relying on a single rule-based mechanism, the module adopts a layered transformation approach that simultaneously considers bibliographic meaning and data structure, thereby ensuring both data consistency and scalability.

4.2. Process stages of the transformation module

The transformation framework proposed in this study consists of a series of modular processing stages, each operating independently while remaining structurally interconnected. The overall transformation process is composed of the following five stages.

(1) Data Normalization stage.

At this stage, input data with heterogeneous structures, such as KORMARC and MODS, are transformed into a unified internal representation. Tasks such as standardizing field structures, harmonizing character encoding, and processing repeatable fields are performed to ensure structural consistency, thereby enabling subsequent semantic analysis.

(2) Semantic Analysis stage.

In this stage, the meaning of each metadata element is analyzed and decomposed into entities, properties, and relationships. For example, the system identifies whether a particular field represents an author, publication information, or a subject term, and defines the functional role of each element accordingly.

(3) Entity Extraction stage.

Based on the results of semantic analysis, bibliographic data are decomposed into core BIBFRAME entities such as Work, Instance, and Item. During this process, identification rules are applied to prevent duplication of entities, and each entity is assigned a unique URI.

(4) Relationship Generation stage.

This stage defines and establishes relationships among the extracted entities. For instance, it explicitly represents authorship relationships between Work and Agent, and

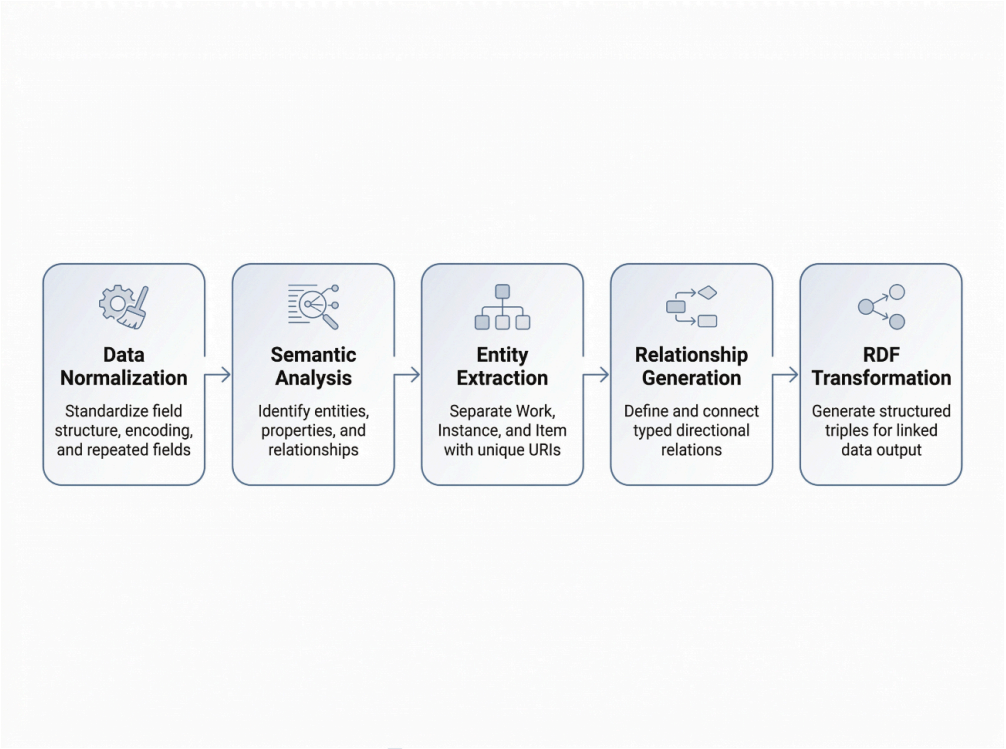

Figure 2. *Process stages of the transformation module based on the normalization*

publication relationships between Instance and publisher entities. It is essential at this stage to clearly specify both the type and direction of relationships.

(5) RDF Transformation stage.

Finally, all entities and relationships are transformed into RDF triple structures for output. In this process, each entity and property is represented by a URI, resulting in data that can be directly utilized in linked data environments.

This modular structure ensures flexibility and scalability in the transformation process, allowing for independent refinement and extension at each stage.

4.3. Transformation rules and processing strategies

The processing stages of the transformation module follow a sequential workflow consisting of data normalization, semantic analysis, entity extraction, relationship generation, and RDF transformation. Each stage plays a role in progressively structuring the semantic content of bibliographic data. To ensure the effective operation of this stepwise process, it is necessary to clearly define the transformation criteria and processing strategies applied at each stage.

To enhance clarity and consistency in the transformation process, this study categorizes bibliographic elements into three types: full transformation, partial transformation, and non-transformable elements.

Full transformation refers to cases where elements can be directly mapped to BIBFRAME classes and properties and converted into RDF triples without structural loss. Partial transformation applies to cases where some degree of semantic reduction

or restructuring is required. Non-transformable elements are those for which no direct correspondence exists within the BIBFRAME structure.

These transformation categories are closely linked to the semantic analysis and relationship generation stages described in Section 4.2. Once the functional role of each bibliographic element is defined during semantic analysis, the corresponding transformation strategy is determined based on its classification. This, in turn, is reflected in the entity extraction and relationship generation stages. In this sense, transformation rules are not confined to a single stage but function as guiding principles throughout the entire transformation process.

Furthermore, this study adopts a hybrid strategy that combines automated processing with expert judgment. Entities with relatively well-defined structures, such as Instance and Item, are processed through rule-based automation, whereas entities requiring semantic interpretation, such as Work, are refined through expert intervention. This approach ensures both processing efficiency and bibliographic accuracy.

Overall, these transformation rules and processing strategies go beyond simple format conversion, serving to coherently integrate the stepwise processing framework into a semantically structured system. As a result, the proposed transformation module implements an integrated framework that reconstructs bibliographic data into an entity- and relationship-based linked data model through the organic combination of processing stages and transformation strategies.

4.4. Strategic Considerations for BIBFRAME transformation

The transformation rules and processing strategies proposed are intended as a conceptual and methodological framework for semantic-based bibliographic transformation rather than as a fully implemented software system. Accordingly, the focus of this study is on defining the principles, procedures, and decision criteria required to reconstruct bibliographic data from record-based structures into an entity–relationship model aligned with BIBFRAME.

From this perspective, the proposed strategies should be understood as guidelines for transformation rather than as implementation-specific procedures. In particular, the classification of bibliographic elements, the normalization process, the identification of entities, and the generation of relationships are presented as semantic design principles that can be adapted to different metadata formats and system environments.

Although the framework was designed with consideration of existing technologies and transformation environments, such as MARC processing tools, RDF generation frameworks, and linked data infrastructures, the actual implementation and validation of a software-based transformation module are beyond the scope of the present study. Future research will focus on developing a prototype system, applying the proposed framework to real KORMARC and MODS datasets, and evaluating transformation accuracy, semantic consistency, and interoperability in practical library environments.

Therefore, the contribution of this study lies not in the implementation of a specific conversion tool, but in establishing a semantic-based transformation strategy that provides a theoretical and methodological foundation for future BIBFRAME transformation systems.

5. Conclusion

This study proposed a semantic-based transformation framework and module design for converting KORMARC- and MODS-based bibliographic data into the BIBFRAME structure. Recognizing the limitations of existing approaches that rely on syntactic field-to-field mapping and fail to adequately capture bibliographic meaning and relationships among entities, this study adopted an approach that reinterprets bibliographic data as units of meaning and reconstructs them into an entity- and relationship-oriented structure.

To achieve this, a bibliographic normalization model was developed to integrate structural differences among heterogeneous metadata formats. Based on this model, a stepwise transformation module consisting of semantic analysis, entity extraction, relationship generation, and RDF transformation was proposed. In particular, by redefining bibliographic elements as functional semantic units and aligning them with the class and property structures of BIBFRAME, this study provides a foundation for transforming record-based data into an entity-based linked data model.

In addition, this study presented transformation rules, processing strategies, and practical implementation methods, demonstrating that the proposed transformation module extends beyond a conceptual model and is applicable in real-world environments. The hybrid transformation strategy, combining automated processing with expert validation, was shown to effectively balance processing efficiency and bibliographic accuracy.

In conclusion, this study re-conceptualizes bibliographic data transformation as a process of semantic structure reconstruction rather than a simple format conversion, and proposes a systematic model and methodology for its implementation. This approach contributes to enhancing the interoperability and reusability of bibliographic data and provides a foundation for extending library data into knowledge graphs within linked data environments.

Future research should focus on applying the proposed framework to large-scale real-world datasets, validating its performance, extending it to diverse resource types, and developing strategies for integrating authority data and external identifiers.

References

[1] Library of Congress, “Bibliographic framework as a web of data: Linked data model and supporting services,” 2012. [Online]. Available: https://www.loc.gov/bibframe/pdf/marcld-report-11-21-2012.pdf

[2] P. Schreur, *The evolution of BIBFRAME: From MARC surrogate to web conformant data model*. in Proceedings of the IFLA World Library and Information Congress. 2018. [Online]. Available: https://repository.ifla.org/rest/api/core/bitstreams/c08af807-923d-4f57-a7e7-0e2aa46dbbe5/content

[3] J. Park, A. Brenza, and L. Richards, "BIBFRAME linked data: A conceptual study on the prevailing content standards and data model," in *Linked Open Data: Applications, Trends and Future Development*, K. Okoye, Ed., IntechOpen, 2020. doi: 10.5772/intechopen.91849.

[4] M. Peponakis, "In the name of the name: RDF literals, ER attributes, and the potential to rethink the structures and visualizations of catalogs," *Information Technology and Libraries*, vol. 35, no. 2, pp. 19–38, 2016.

[5] P. Gaitanou, E. Garoufallou, and M. Zervas, "Linked data for libraries: Creating a global knowledge space, a systematic literature review," *Journal of Information Science*, vol. 50, no. 1, 2024, doi: 10.1177/01655515221084645.

[6] J. Hyatt and J. Godby, "Comparison of key entities within bibliographic conceptual models and implementations," *Library Resources & Technical Services*, vol. 67, no. 1, 2023, [Online]. Available: https://journals.ala.org/index.php/lrts/article/view/7345/10100

[7] M. Guerrini and T. Possemato, "From record management to data management: RDA and new application models BIBFRAME, RIMMF, and OliSuite/WeCat," *Cataloging & Classification Quarterly*, vol. 54, no. 3, pp. 179–199, 2016, doi: 10.1080/01639374.2016.1144667.

[8] Q. Jin, J. Hahn, and G. Corll, "BIBFRAME transformation for enhanced discovery," *Library Resources & Technical Services*, vol. 60, no. 4, pp. 223–245, 2016, doi: 10.5860/lrts.60n4.223.

[9] M. Sfakakis, S. Zapounidou, and C. Papatheodorou, "Mapping derivative relationships from BIBFRAME 2.0 to RDA," *Cataloging & Classification Quarterly*, vol. 58, no. 7, pp. 603–631, 2020, doi: 10.1080/01639374.2020.1821856.

[10] S. Zapounidou, M. Sfakakis, and C. Papatheodorou, "Mapping derivative relationships from RDA to BIBFRAME 2," *Cataloging & Classification Quarterly*, vol. 57, no. 5, pp. 278–308, 2019, doi: 10.1080/01639374.2019.1650152.

[11] C. J. Godby, "The relationship between BIBFRAME and OCLC's linked-data model of bibliographic description: A working paper," OCLC Research, 2013. [Online]. Available: https://www.oclc.org/content/dam/research/publications/library/2013/2013-05.pdf

[12] B. M. Gonzales, "Linking libraries to the web: Linked data and the future of the bibliographic record," *Information Technology and Libraries*, vol. 33, pp. 10–22, 2014, doi: 10.6017/ital.v33i4.5631.

[13] S. e. a. Hayat, "A systematic review of linked data and RDF adoption in libraries: Benefits, challenges and best practices," *Advance Social Science Archive Journal*, vol. 4, no. 1, pp. 2180–2192, 2025, [Online]. Available: https://assajournal.com/index.php/36/article/view/706/1026

[14] B. Michael and M. K. Han, *Assessing BIBFRAME 2.0: Exploratory implementation in Metadata Maker*. in Proceedings of the 2019 International Conference on Dublin Core and Metadata Applications. 2019, pp. 26–31. doi: 10.23106/dcmi.952141614.

[15] J. Samples and I. Bigelow, "MARC to BIBFRAME: Converting the PCC to linked data," *Cataloging & Classification Quarterly*, vol. 58, no. 3–4, pp. 403–417, 2020, doi: 10.1080/01639374.2020.1751764.