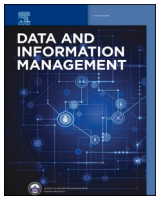
9 (2025) 100068

Contents lists available at ScienceDirect

Data and Information Management

journal homepage: www.journals.elsevier.com/data-and-information-management



Metadata application profile as a mechanism for semantic interoperability in FAIR and open data publishing

Nishad Thalhath *, Mitsuharu Nagamori, Tetsuo Sakaguchi

School of Library, Information and Media Studies, University of Tsukuba, Tsukuba, Japan

ARTICLE INFO

Keywords:
Application profiles
Semantic interoperability
Data Interoperability
Open data
FAIR data
Linking data
Semantic validation
Linked Open Data
RDF

ABSTRACT

Application profiles, also known as metadata application profiles, are customised collections of vocabularies adapted from various namespaces and tailored for specific local applications. These profiles act as constrainers and explainers for the (meta)data. Semantic interoperability is the ability of computer systems to exchange data in a mutually understandable manner, facilitating data sharing across diverse platforms and applications without compromising its meaning. As a critical component of semantic interoperability, application profiles enforce semantics to (meta)data, enhancing its openness, interoperability, and reusability. This study assesses the feasibility of representing a comprehensive application profile in a format aligned with the semantic web, ensuring interoperability between profiles and datasets. Dublin Core Description Set Profiles (DSP) is adapted as the modeling framework for metadata application profiles, steering the associated datasets toward RDF compliance. The research outcomes include "Yet Another Metadata Application Profiles" (YAMA) as a preprocessor grounded in the DSP framework for developing and managing metadata application profiles. YAMA facilitates the generation of various standard formats of application profiles, ensuring they are represented in human-readable documentation, machine-actionable forms, and even data validation languages. A data mapping extension to YAMA is proposed to ensure the semantic interoperability of open data, bridging non-RDF data structures to RDF, thus enabling the publication of 5-star open data. This ensures smooth dataset integration and the creation of linkable, semantically rich open datasets. The work emphasizes the pivotal role of application profiles in fortifying the semantic interoperability of (meta)data, thereby elevating dataset openness.

1. Introduction

In an era where data is paramount, there is an escalating need to ensure it is useable, reusable, and meaningful. As data production grows exponentially, strategies ensuring efficient, trustworthy, sustainable, and interoperable usage are crucial (George et al., 2014). Interoperability pertains to the capability of systems to effectively exchange and utilise information, becoming especially pivotal when diverse data producers and consumers operating on disparate, incompatible systems need to share information. In contrast, semantic interoperability describes systems' ability to exchange data in a manner meaningful to both parties. Historically, the semantics of data, vital for understanding its meaning, were conveyed through metadata or annotations. While web technologies and open data practices have transformed data dissemination and accessibility, they often compromise data's inherent structure and semantics. However, semantic web technologies like the Resource Description Framework (RDF) and the Web Ontology Language (OWL) adeptly embed data's semantics, enriching its inherent value.

Open data is vital in the contemporary world of data-oriented processes and progress. This research delves into the potential of leveraging metadata application profiles to foster semantic interoperability among open datasets. A primary insight guiding this investigation is the understanding that the mere exchange of datasets is insufficient for open data publishing; the semantics encapsulated within these datasets is equally pivotal. It is essential to acknowledge that open data loses its essence of openness if the metadata and semantics accompanying the (meta)data remain cloistered (Bestek et al., 2022). Consequently, this study aspires to engineer a robust mechanism that meticulously documents datasets' semantics and structural intricacies in formats amenable to human comprehension and machine processing. Building on the foundational principles of application profiles, this research highlights their role as theoretical constructs and pragmatic instruments to ensure semantic transparency and interoperability. This approach ensures

* Corresponding author.
E-mail addresses: nishad@slis.tsukuba.ac.jp (N. Thalhath), nagamori@slis.tsukuba.ac.jp (M. Nagamori), saka@slis.tsukuba.ac.jp (T. Sakaguchi).

https://doi.org/10.1016/j.dim.2024.100068
Received 30 April 2023; Received in revised form 27 November 2023; Accepted 8 February 2024
Available online 22 February 2024

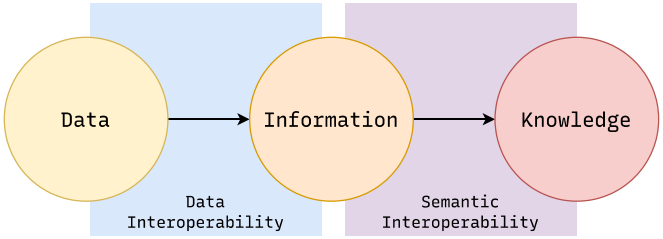


Fig. 1. Role of semantic interoperability in DIK continuum.

alignment with prevalent frameworks and best practices guiding open data publication.

This research makes a noteworthy contribution to the field of data publishing by introducing a simplified framework aimed at representing semantic interoperability within the realms of FAIR and open data. Central to this study is the exploration of a practical approach, grounded in existing proposals, to develop comprehensive, semantically-aware profiles for datasets and endpoints. The project capitalizes on the capabilities of semantic web technologies, enhancing data openness by ensuring the openness of semantics, documentation, and validation processes. Additionally, the study introduces a simplified format coupled with user-friendly proof-of-concept tools, facilitating effortless achievement of semantic interoperability in the publishing of open data. The importance of semantic interoperability in improving the usability, sharing, and overall quality of open data is a key focus. Building upon and refining the concept of application profiles, the framework presented is both adaptable and accessible, making it suitable for a wide range of domains. This versatility significantly extends the framework's practicality and influence in the spheres of open and FAIR data publishing and use.

2. Literature review

2.1. Significance of semantic interoperability

The Data-Information-Knowledge-Wisdom (DIKW) continuum illuminates the journey from raw, uninterpreted data to the application of real-world knowledge (Rowley, 2007). This progression begins with raw data, which is then processed into actionable information. Further refinement and interpretation transform this information into knowledge, which aids understanding and decision-making. The pinnacle, wisdom, represents the practical application of this knowledge in addressing real-world challenges. In this context, basic interoperability, or the ability of systems to exchange information effectively, is foundational (OECD, 2021). However, semantic interoperability becomes indispensable to truly elevate information into knowledge, as illustrated in Fig. 1. Achieving semantic interoperability requires standards defining the meaning of data and ensuring accurate interpretation of this meaning across varied systems.

2.2. FAIR data and open data

The FAIR data principles advocate for data to be Findable, Accessible, Interoperable, and Reusable, facilitating its use by humans and machines. Formulated by international experts spanning data science, data management, and scholarly communication, these guidelines are adaptable across various disciplines and datasets rather than being rigid rules (Wilkinson, 2016). They benefit data providers, users, and stewards alike. These principles can revolutionise data management by enhancing its utility and value by endorsing findability, accessibility, interoperability, and reusability. Central tenets encompass findability via open standards and searchable indices, accessibility using standardized methods, interoperability amongst diverse systems, and reusability strengthened by open licensing and comprehensive documentation. These standards define and ensure "value" and "valuable data" for in the domain of reusable data publishing (Bezuidenhout, 2020).

FAIR compliance necessitates a clear and concise semantic model to depict the dataset's entities and relationships. Constructing an effective semantic model, especially considering dataset complexities, remains important. This model should echo a consensus perspective pertaining to a specific domain and intent. Leveraging pre-existing ontologies and vocabularies can alleviate the modeling process (Wilkinson et al., 2017). While the FAIR principles vouch for open data's accessibility, reusability, and interoperability, they concurrently stress its uninhibited availability, exempting privacy or ethical reservations. It's crucial to differentiate between "FAIR data" and "open data". The former underscores findability, accessibility, interoperability, and reusability, while the latter implies data that are freely accessible and devoid of any constraints. Not all open data adheres to FAIR standards, and not all FAIR data is inherently open due to potential licensing, privacy, or ethical considerations. Therefore, the FAIR principles emphasize the importance of well-documented and appropriately characterized data that can be used for various applications.

2.3. RDF, linked open data, 5-star open data, and 7-star open data

RDF, a standard for representing information on the Semantic Web, an extension of the World Wide Web, is expressed in triples consisting of a subject, predicate, and object (Wood et al., 2014). This framework and its related specifications were introduced and maintained by the World Wide Web Consortium (W3C). RDF data structure can be expressed in several formats, including RDF/XML, RDFa, JSON-LD, and Turtle. The directed graph model RDF uses sets it apart from other data formats and provides greater flexibility and power in data representation (Bizer et al., 2011). In a graph-based format, data is represented as interconnected nodes in a graph instead of as a table or set of key-value pairs (Wood et al., 2014). The standardisation of RDF also facilitates the exchange and interoperability of data and querying data using a query language called SPARQL (SPARQL Protocol and RDF Query Language) (Seaborne and Harris, 2013).

Linked Open Data is a method of data publishing on the web in a way that it can be interlinked. It is based on the idea of linking data on the Semantic Web using RDF. Linked Data is a way to create a network of data by connecting data from different sources. The RDF standard enables linked data by allowing various data sources to be linked together. Linked Data aims to create a more interconnected and interoperable web of data (Wilkinson et al., 2017). The four principles of Linked Data are: 1) Use URIs as identifiers for things, 2) Use HTTP URIs so that users can access those resources, 3) When a user agent accesses a URI, it should provide useful information using standards such as RDF, and 4) Include links to other URIs so that they can connect to more things. LOD explicitly mandates data openness, while FAIR emphasizes the need for a defined license for access, incorporating reusability within the scope of the license agreement. Additionally, FAIR significantly focuses on the necessary contextual information, such as provenance details, to enhance data reuse. While LOD principles also regard such metadata as interoperable, the emphasis FAIR places on augmenting data with metadata suggests that it builds upon and extends the LOD framework (Hasnain et al., 2018).

Tim Berners-Lee proposed a five-star rating system to evaluate the openness and usability of data on the web (Berners-Lee, 2006). At its most basic, a single star signifies data that is available on the web with an open license, thus qualifying as Open Data. A two-star rating indicates a more accessible tier, with data available in machine-readable structured formats, like an Excel spreadsheet, as opposed to a mere image scan of a table. Advancing further, three-star data is not only machine-readable but also presented in a non-proprietary format such as CSV, making it more universally accessible than proprietary formats like Excel. The four-star level goes beyond accessibility and delves into

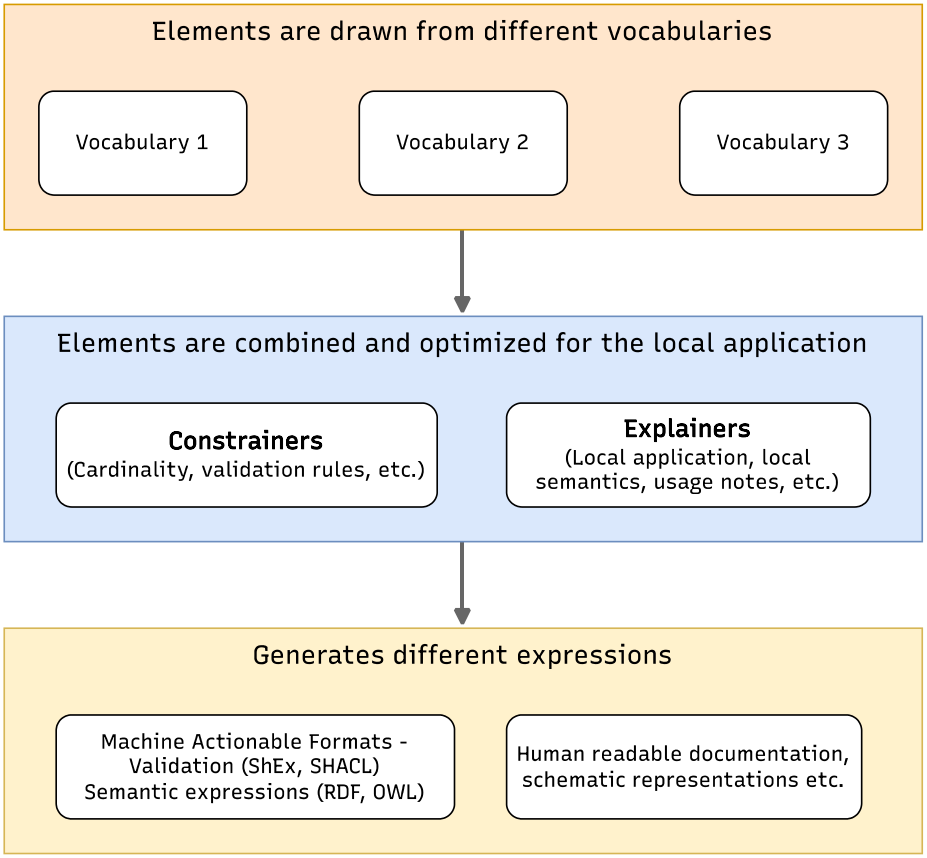


Fig. 2. Overview of an application profile.

standardized identification, urging the use of W3C's open standards, specifically RDF and SPARQL. This enables a universal reference system, allowing external entities to reference or "point at" the data. The pinnacle, a five-star rating, encapsulates all the prior tiers and additionally emphasizes the importance of data interconnectedness. Here, data providers are encouraged to link their datasets with others, enriching them with contextual relevance and fostering a more integrated web of data. For open data publishers, using the RDF format can ensure this higher level of openness and interoperability.

A significant challenge of dataset re-using is evaluating the data's suitability for the local application. Datasets that are not adequately documented and the vocabularies used make them less reusable for the local application of the data. Also, even if the schema is documented, evaluating how well the data aligns with the expected schema may be difficult and ensure the data quality without an actionable validation mechanism (Hyvönen et al., 2014).

To address these issues, Hyvönen et al. (Hyvönen et al., 2014) proposed two extra stars.

- The 6th star is given if the schemas (vocabularies) used in the dataset are explicitly described and published alongside the dataset unless the schemas are already available somewhere on the Web.
- For the 7th star, the quality of the dataset against the schemas used in it must be explicated so that the user can evaluate whether the data quality matches her needs.

2.4. Data profiles and semantic profiles for open data

Profiles in open data play an indispensable role in ensuring datasets are comprehensive, meticulously detailed, and user-friendly. Such profiles lay down guidelines for articulating the data's structure, essence, and backdrop, simplifying the data comprehension and its integration potential with other platforms. An example in this context is the Frictionless Data Package (Walsh and Pollock). This lightweight descriptor details datasets along with pertinent metadata, encapsulating data structure, schema, provenance, and licensing intricacies. Other notable data profiles are the Data Catalog Vocabulary (DCAT) (Albertoni et al)– standardising data catalogue description, and the Data Documentation Initiative (DDI) (DDI Lifecycle 3.3) – delineating survey data's content, structure, and context. Employing such profiles ensures that datasets are thoroughly documented, systematically detailed, and primed for easy utilization.

In comparing data and semantic profiles, despite their shared objective of elucidating datasets and related metadata, distinct differences emerge crucial for data openness and cross-compatibility. While data profiles predominantly focus on the content, structure, and setting of data, semantic profiles traverse further. They express the data's semantics, harnessing technologies like ontologies and linked data to offer a machine-interpretable, detailed description of (meta)data, casting light on data elements' interrelations and their intrinsic semantic significance. This characteristic makes semantic profiles a suitable means for semantic interoperability.

Both these profiles – data and semantic, are foundational for data's openness and interoperability. While the former fortifies data interoperability, the latter is pivotal for semantic interoperability. Initiatives aiming at semantic-aware data profiles have given rise to methodologies like MetaProfiles (Thalhath, Nagamori, & Sakaguchi, 2020) and vocabularies such as the Profiles Vocabulary (Car, 2019), which aspire to craft unified data profiles.

2.5. Application profiles, DCAP and DSP

Application profiles, often called Metadata Application Profiles, are a combination of vocabularies, which are mixed and matched from

Table 1
The related works successfully meet the requirements of Application Profile Expressions.

Type	Related Works
Human Readable	Simple-DSP, DCTAP
Machine Actionable	OWL-DSP, Simple-DSP, DCTAP
Hybrid (Both Machine and Human Friendly)	Simple-DSP, DC-TAP
Validation Formats	ShEx, SHACL

different namespaces and optimised for a particular local application. Application profiles express the terms taken from other namespaces and the structural use of those terms in the local instance data (Heery & Patel, 2000). Application profiles also express constraints on those terms so that the data can be validated as well. An overview of a typical application profile is illustrated in Fig. 2.

A Dublin Core Application Profile (DCAP) specifies how some metadata description sets are constructed. It includes information on the terms used in the description sets, how they are deployed, and constraints on the values and datatypes of the properties used (Nilsson et al). The Singapore Framework for Dublin Core Application Profiles is a set of standards for designing metadata application profiles that are interoperable and reusable. Singapore Framework ensures that such application profiles maintain documentary completeness and align seamlessly with Web-architectural principles (Nilsson et al). Description Set Profiles (DSP) is a Dublin Core Application Profiles constraint language. DSP is based on the DCMI Abstract Model (DCAM), which defines Description Set, Description, and Statement. A DSP defines constraints on Description Sets, Descriptions, and Statements. Description Set Templates hold one or more Description Templates composed of Statement Constraints. DSP supports the RDF-oriented data design with properties and datatypes (Nilsson).

2.6. Existing approaches to enhance semantic interoperability in FAIR and open data

Various approaches have been explored to address the challenge of semantic interoperability in FAIR and open data. Meredith et al. proposes aligning semantic interoperability frameworks with the FOXS stack specifically for the health data (Meredith et al., 2022). The FAIREr (FAIR + human Explorability raised) model emphasized human explorability and cognitive interoperability aspects to ensure the semantics in the FAIR data interoperability (Vogt & Extending FAIR to FAIREr, 2023). For data model unification in FAIR, Stupnikov and Kalinichenko examine techniques for harmonizing data models (Stupnikov & Kalinichenko, 2019). Wilkinson et al. investigate the application of Web technologies in data discovery and integration (Wilkinson et al., 2017). The notion of semantic values, introduced by Sciore et al. serves as a tool to enhance semantic interoperability in the FAIR data (Sciore et al., 1994). Strawn proposes FAIR Digital Objects as a very general virtual layer to stack on top of data systems to overcome data interoperability obstacles for heterogeneous data (Strawn, 2019).

Focusing on open data, Davies underscores the importance of metadata standards in achieving semantic interoperability, particularly in the electronic governance (Davies et al., 2008). Lin explores the quality of government open data portals, shedding light on their efficiency and user-friendliness (Cathy, 2018). Similarly, Hoxha and Brahaj advocate for a semantic approach in publishing and visualizing open government data, focusing on improving data usability and interpretation (Hoxha & Brahaj, 2011). Meanwhile, Gal as well as Amato et al. address the challenges of semantic interoperability in distributed data sources (Gal, 1999) (Amato et al., 2013). Gal emphasizes the need for sophisticated technologies to guarantee high-quality, semantic-rich open data. Amato et al. introduce a cloud-based framework designed to support semantic interoperability in open data. In the field of geospatial open data, Paul and Ghosh suggest a new methodology for interoperable access with the semantics (Paul & Ghosh, 2012), whereas Harvey et al. stress the importance of resolving semantic differences to facilitate effective sharing of geographic information (Harvey et al., 1999).

Semantic web technologies are becoming well-adapted as a means of semantic interoperability in FAIR and open data. Earth science ontologies are proposed to overcome terminological discrepancies and enhance data interoperability (Raskin et al., 2004). Linked data practices were explored in the FAIR digital objects proposal (Soiland-Reyes et al., 2022) and rely on semantic web and linked open data principles. Various semantic web standards were tested and accepted for publishing and integrating open data (Polleres & Steyskal, 2015). This acceptance of semantic web in publishing FAIR and open data further highlights the significance of this study in which a semantic web-oriented approach of application profiles is adapted as mechanism for extending semantic interoperability.

2.7. Application profile and RDF mapping state-of-the-art

There are different attempts for data profiling and RDF mapping languages. A brief overview of the state-of-the-art is provided below.

DC Tabular Application Profiles (DC-TAP) methodology facilitates the generation of application profiles using a tabular format. These tables are capable of being stored in a CSV file structure, thereby rendering them comprehensible to both computer programs and human readers (Coyle et al., 2023).

ShEx (Shape Expressions) is a language for expressing constraints on RDF (Resource Description Framework) graphs. It is used to validate RDF graphs against a set of rules or constraints, which helps ensure the data in the graph is accurate and consistent (Thornton et al. et al., 2019).

SHACL (Shapes Constraint Language) is a standard established by W3C for checking the integrity of an RDF-based graph database. It delineates how validation outcomes should be presented to provide users with insightful alerts (Kontokostas & Knublauch, 2017).

Tarql: SPARQL for Tables is a command-line tool that uses SPARQL 1.1 syntax to convert CSV files to RDF (Cyganiak, 2015).

LinkML is a flexible modeling language that allows authors to create schemas in YAML that describe the structure of data. LinkML is also a framework for working with and validating data in a variety of formats (JSON, RDF, TSV) and can be used to compile LinkML schemas to other frameworks (Moxon et al., 2021, pp. 148–151).

R2RML is a language expressing customized mappings from relational databases to RDF datasets. This language allows different mapping implementations, such as creating a virtual SPARQL endpoint over the mapped relational data, generating RDF dumps, or offering a Linked Data interface (Das et al., 2012).

RDF Mapping Language (RML) is a mapping language that can express customized mapping rules from heterogeneous data structures and serializations to the RDF data model. RML is defined as a superset of the W3C-standardized mapping language R2RML (Dimou et al., 2014).

YARRRML is a human-readable text-based representation for declarative Linked Data generation rules. It is a subset of YAML that can be used to represent R2RML and RML rules (Van Assche et al., 2021).

CSV2RDF defines the procedures and rules for converting tabular data into RDF, including how metadata annotations can describe the structure, meaning, and interrelation of tabular data (Tandy et al., 2015).

RDF Transform is an extension for OpenRefine that allows users to transform data into RDF formats. The RDF Transform extension provides a graphical user interface (GUI) for transforming OpenRefine project data to RDF-based formats. The RDF transform extension maps the data with a template graph designed using the GUI (GitHub - AtesComp et al).

OWL-DSP and Simple-DSP are part of the MetaBridge project. OWL-DSP is a scheme that uses Web Ontology Language (OWL) to define description set profiles (DSP) for the description of an application profile, with all schemas in this system represented in RDF and OWL. SimpleDSP is a DSP representation based on a tabular format (Nagamori

Table 2
Related works in satisfying 6- and 7-star open data model.

Star	Application Profile components	Related Works
6	Human-readable documentation, Machine actionable formats	Documentation, OWL-DSP, Simple-DSP, DCTAP
7	6 Star + Profiles expressed in validation formats	ShEx, SHACL

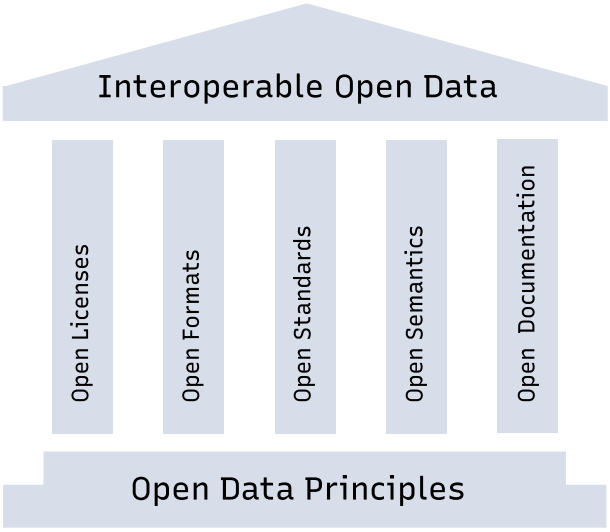


Fig. 3. 5-Pillars model for interoperable open data.

et al., 2011).

In order to assess the current landscape and the alignment of related projects with appropriate application profiles, we provide an overview in Table 1. This table outlines the state-of-the-art projects and their compatibility with a proper application profile. Additionally, Table 2 illustrates the extent to which these projects adhere to the 6 and 7-star data model concepts. This information is vital for understanding the current practices in the field and identifying areas for improvement or further alignment with these concepts.

From the survey on related works, it is learned that there's a pressing need to incorporate a pre-processor in creating metadata application profiles, given its diverse range of applications. Firstly, pre-processors are pivotal in crafting practical Metadata Application Profiles. Furthermore, they enhance the sustainability of these profiles by providing a foundational format. Notably, using a structured textual format is ideal in collaborative development settings and version control systems, facilitating straightforward change tracking, whether through elementary diff operations or more intricate continuous integration systems. Additionally, the structured nature of a pre-processor format ensures that the validation of metadata application profiles is streamlined, thereby minimising errors and potential logical complications. Lastly, the inclusion of optional change records in a pre-processor format acts as a developmental guide (Thalhath, Nagamori, Sakaguchi, et al., 2020). This facilitates the recreation of older versions, the production of easily understandable changelogs, and the formulation of formats beneficial for metadata crosswalks and migrations.

3. Theoretical frameworks

3.1. Combining the 7-star open data and the FAIRness of data

Combining the FAIR principles with the 7-star open data model enhances semantic interoperability by ensuring that data is open, accessible, and thoroughly documented. The FAIR principles emphasize the importance of clear metadata and the adoption of shared vocabularies, which is crucial for achieving semantic interoperability. On the other hand, the 7-star open data model provides a framework for assessing data quality and appropriateness, ensuring alignment with intended purposes and adherence to relevant standards. By integrating these frameworks, data producers and consumers can collaboratively create high-quality, well-documented, and interoperable data with shared semantics, which can be easily shared and used across various systems, unlocking open data's full potential.

Publishing application profiles is vital for achieving a 6-star level of data openness. Embedding these profiles within the RDF data improves interoperability and reusability. According to the FAIR data principles, data interoperability involves integrating data with other data and ensuring compatibility with analysis, storage, and processing applications or workflows. A 6-star level of data is highly interoperable and reusable. During the FAIRification process, defining a semantic model for the dataset is crucial. This model describes the meaning of entities and relationships in the dataset in a machine-readable format. Application profiles can serve as localised ontologies for the data they represent, profiling the application of terms, concepts, and structures.

For a 7-star level, data quality must be ensured. Providing a validation schema allows users to assess data quality. By declaring constraints in an application profile, machine-readable validation schemas or formats can be inferred, ensuring data quality.

3.2. 5-Pillars Model for Interoperable Open Data

FAIR principles provide a framework for ensuring data interoperability but are not strictly adhered to open data. This research proposes a 5-pillar model to ensure complete interoperability of open data, with a strict foundation on open data principles (ODC). The 5 pillars are represented in Fig. 3.

1. **Open licenses** - Open licenses primarily address the legal rights for data reuse rather than the technical aspects of interoperability. However, the legal aspects intertwined with these licenses play a significant role in data interoperability. For instance, a dataset that is technically interoperable can still be rendered unusable if its license doesn't permit integration with other datasets. It's essential to choose open licenses that support both the technical and legal aspects of interoperability. For example, using a widely accepted open license like the Creative Commons Attribution (CC BY) ensures that the data is not only useable on its own but also can be legally combined or "stacked" with datasets under other licenses. This ensures a harmonious flow of data integration and reduces legal barriers.[1]
2. **Open Formats** - Open data/file formats play a pivotal role in ensuring smooth data interoperability. Such formats are non-proprietary and devoid of restrictive intellectual property rights, which means they can be freely used, shared, and modified by anyone. Embracing open file formats ensures that open data can seamlessly move across diverse platforms and applications, sidestepping the hurdles of complex data conversions or reliance on proprietary software. For instance, consider the CSV (Comma-Separated Values) format, an open format widely used for data storage. Because CSV is open and universal, data stored as a CSV file can be effortlessly imported into various applications, from spreadsheet software like Microsoft Excel to databases like PostgreSQL, without the loss or distortion of information. Moreover, open formats like CSV naturally align with other similar formats, promoting interoperability with minimal fuss and expense.
3. **Open Standards** - Open standards are publicly available formal technical specifications developed and maintained via a collaborative and consensus-driven process (ITU-T, 2005). They define how

[1] https://mozillascience.github.io/open-data-primers/5.3-license-stacking.html.

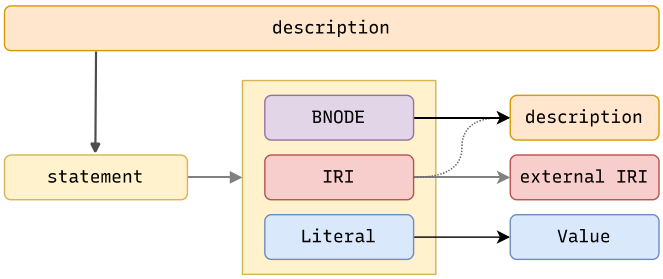


Fig. 4. Structure of a DSP-based application profile.

data should be structured, formatted, and exchanged. Adhering to open standards allows different systems and applications to interoperate seamlessly, reducing the need for custom integrations, tooling, and complex data conversions. For instance, the Extensible Markup Language (XML), a flexible, text-based format designated to store and transport data, is an open standard that exemplifies how diverse systems can achieve interoperability.

4. **Open Semantics** - To ensure effective use of open data and interdisciplinary interoperability, it's important to use open and accessible vocabularies and ontologies. These allow for a shared understanding of the meaning of data and enable open semantic interoperability. Creating application profiles from open vocabularies, ontologies, and public knowledge graphs ensures that the data is semantically open and interoperable. This also ensures that not only the data, but the semantics of the data are open. For instance, when medical researchers from various countries join a global health initiative, employing an open ontology, such as "Disease Ontology", guarantees consistent interpretation of disease data irrespective of their diverse backgrounds. Moreover, given that Disease Ontology is licensed under CC0, users of this open data can freely discern its semantics (Schriml et al., 2023), and the data can be easily mapped with many open knowledge graphs such as Wikidata (Thalhath et al., 2021).
5. **Open Documentation** - For data to be truly interoperable, its documentation must be thorough, detailing the data and its underlying semantics. This documentation should be openly licensed, easy to access, and presented in an open format. Essential elements of this documentation include the data's structure, meaning (semantics), origin (provenance), illustrative examples, and licensing details. Offering this documentation in both human-friendly and machine-readable formats aids in manual data interpretation and streamlines automated data processes. For example, a city's public transportation data might contain detailed documentation describing each field, like bus numbers, routes, and timings. If this documentation is also machine-readable, app developers can easily integrate it into a transit app, guiding users on their daily commutes. Robustly constructed semantic documentation has the potential to produce linked open data, facilitating data linking with established open knowledge bases, such as the OpenStreetMap.[2]

Based on this model, this research considers the 4th and 5th pillars as the core of semantics interoperability for open data, and application profiles serve a crucial role in documenting semantics to strengthen the 5th pillar.

3.3. YAML as an authoring format for application profiles

YAML Ain't Markup Language (YAML) is a powerful and user-friendly data serialisation standard. Its adaptability is reflected in its compatibility across the most prevalent programming languages. According to the recent specification v1.2, YAML is recognised as a superset of JSON (Ben-Kiki et al). Its clear emphasis on readability renders YAML a preferable choice for manual data serialisation and modifications, especially when compared to formats like CSV. Its compatibility with version control systems, such as Git, and text editors further underscores its utility. Its open-format nature wards off vendor lock-in scenarios for documents, fostering a suitable environment for methodical and systemic programmatic interactions with YAML documents. YAML's flexibility is evidenced in its adoption in initiatives like the OpenAPI Specification, which standardises RESTful API interfaces (OpenAPI Initiative), YARRRML, a textual representation that sets out Linked Data generation rules (Heyvaert et al., 2018), and Dead simple OWL design patterns (DOS-DP), a streamlined system specifying OWL class patterns (Osumi-Sutherland et al., 2017). Application profiles' structured and logically lucid nature finds a conducive medium in YAML. When expressed in YAML, application profiles are seamlessly structured, sidestepping any convoluted processing. Some of YAML's notable attributes encompass its capacity for comments, syntax highlighting, and formatting, all facilitated by modern text editors. This aids in enhancing the visual clarity and organisation of an application profile (Thalhath et al., 2019a).

3.4. Adapting DSP as a framework for application profile

Description Set Profiles (DSP) is an RDF-based application profile framework that needs four distinguished elements to express the profiles.

1. **Descriptions** are the basic representation of a shape or class in the data it represents. Descriptions are a set of statements that describes the entity it represents. In a typical RDF expression of data, descriptions can be inferred as `rdfs:Class`, and can be a blank node (BNODE) as well.
2. **Statements** are the basic triple representation, with a subject and an object with a predicate connecting them. When statements form a description, they will have a common subject and thus become parts of a shape. An object in a statement can be a literal string or linking to an IRI or blank node (BNODE). Constraints are the elements that constrain the statements in shape or values in statements. There are two types of constrainers: statement constrainers and value constrainers.
3. **Statement constraints** are applicable for occurrence rules, such as the rules on mandatory, repeatable, or occurrences of a statement. The type constrainers for statements are another type of statement constrainer. Statement types can be constrained to literal or non-literal such as IRI and BNODE.
4. **Value constraints** are the rules that constrain the value of an object in a statement. In general, value constrainers are more applicable to literal values. However, IRIs can be constrained with a pattern or a specific IRI string.

These elements and their relations are illustrated in Fig. 4.

4. Methodology

4.1. Identifying application profile requirements and use-cases

This study examined two high-level application profiles from prominent agencies to determine fundamental application requirements and modeling concepts.

Firstly, the Digital Public Library of America (DPLA) Metadata Application Profile ,[3] which organizes and validates metadata within the DPLA. Based on the Europeana Data Model (EDM), this profile is tailored to aggregate metadata from American cultural heritage institutions,

[2] https://www.openstreetmap.org.

[3] https://pro.dp.la/hubs/metadata-application-profile.

guiding the storage, serialisation, and dissemination of metadata. Secondly, the DCAT Application Profile for Data Portals in Europe (DCAT-AP) ,[4] a framework based on the Data Catalogue Vocabulary (DCAT), focuses on cataloging European public sector datasets. Its primary function is to facilitate cross-portal searches for datasets, enhancing the findability of open data across various sectors. This profile is designed to ensure open and public data interoperability for diverse use cases.

We analyzed different versions of these application profiles to pinpoint general application profile requirements. Some general-purpose application profiles collected as part of the MetaBridge project[5] are also reviewed. This analysis yielded a comprehensive requirements roadmap, forming the basis for our proposed application profile format. Additionally, we identified various use cases for elements that can be represented and edge cases with limited applicability.

4.2. Selection of Dublin Core description set profiles (DSP) as the modeling framework

The rationale for choosing the Description Set Profiles (DSP) from the Dublin Core Metadata Initiative (DCMI) in this study is anchored in DCMI's long-standing history of establishing metadata standards and advocating for optimal practices in metadata application profiles. The factors influencing this selection include.

1. **Maturity**: DSP is a long-existing, constrained language proposal for Dublin Core Application Profiles, incorporating the Singapore Framework for application profiles and the Dublin Core Abstract Model (Nilsson). DSP includes essential semantics and syntaxes crucial for implementing application profiles tailored for web-based data publishing (Enoksson). This inclusion makes it a thoroughly documented and robust foundation for the current proposal, offering a well-structured starting point that aligns with the requirements of online data dissemination.
2. **Flexibility**: DSP's design is conducive to modular metadata description, making it versatile for various domains and purposes. It provides a straightforward foundation for metadata applications, especially those centered around Dublin Core elements.
3. **Semantic Richness**: Inherently supporting semantic concepts, DSP is constructed using RDF as the core data description model, rendering it a fitting framework for this research.

The decision to adopt DSP was informed by a comparative analysis of various metadata application profile frameworks, focusing on aspects such as expressivity, ease of adoption, and overall maturity. DSP consistently emerged as the most suitable choice throughout this evaluation, excelling in these critical parameters (Thalhath et al., 2019b).

4.3. Development of the application profile format based on the DSP framework

The development of the Application Profile format in this study was structured into four distinct phases.

1. **Phase 1: Design** - The initial phase involved drafting design sketches for the Application Profile format, focusing strongly on human readability. These sketches outlined the elements of the Application Profile, their interrelationships, and their overall scope. Use cases and requirements identified from the sample analysis are used to make the design fit well with the actual requirements.
2. **Phase 2: Prototype Development** - A draft for the application profile and a prototype preprocessor were developed using the initial outlines. This stage was crucial for creating machine-actionable semantic structures to generate common output formats, as the input format is aligned with the Description Set Profile (DSP) framework.
3. **Phase 3: Iteration** - The project moved into an iterative phase, incorporating feedback loops. This process allowed the format to undergo multiple refinements, ensuring it adhered closely to DSP standards while retaining its unique features.
4. **Phase 4: Evaluation** - The final phase involved applying the newly developed format to recreate the Digital Public Library of America Application Profile (DPLA-AP) and the Data Catalogue Vocabulary Application Profile for Data Portals in Europe (DCAT-AP). This step was critical to verify that the format could effectively represent real-world use cases.

Throughout the development process, user-centric design principles were a key focus to guarantee the format's human readability. Preliminary user testing involving mock-ups was conducted to validate this aspect.

4.4. Data mapping procedure & extension to the application profile format

The challenge was to bridge non-RDF data structures to RDF seamlessly.

- Phase 1: Data Inventory - Initial datasets, mainly in CSV formats, were identified. These were sourced from public repositories and open datasets.
- Phase 2: Data Transformation - Custom scripts using Python's RDFLib and PyShEx were developed to map the data into RDF structures and validate. Then, this mapping is ported as an extension guideline for the previously developed application profile format. The mapping logic was modularized for extensibility.
- Phase 3: Mapping Extension - Based on the peculiarities of the identified mapping requirements, the profile format was extended to encompass additional constructs that facilitate easier data bridging to RDF, a richer semantic format.

4.5. Validation and testing mechanisms

A multifaceted approach was adopted for conducting tests and evaluation of the results.

- **Semantic Validity Testing**: SPARQL queries were executed against mapped RDF datasets to ascertain accurate semantic translation.
- **5-Star Open Data Publication Evaluation**: A check based on the 5-star open data criteria [6] was used. Every dataset processed using the YAMA preprocessor was evaluated against these checks and ensured a minimum 4-star openness.
- **Usability Testing and Evaluation:** We evaluated the human readability of the authored application profiles in this proposed format and qualitatively compared with similar state of the art proposals.
- **RDF Testing**: RDFUnit was leveraged for RDF validation, ensuring data integrity post-mapping as well as the RDF representation of the application profiles.

The study utilized several tools and software for data processing and validation: Python's RDFLib for data transformation, PyShEx for ShEx validation, RDFUnit for RDF validation, and Oxigraph for triplestore management and SPARQL validation. Candidate datasets were sourced from various public repositories, primarily focusing on, but not limited to, CSV datasets.

[4] https://op.europa.eu/en/web/eu-vocabularies/dcat-ap.
[5] https://metabridge.jp/infolib/metabridge/menu/?lang=en.
[6] https://5stardata.info/en/.

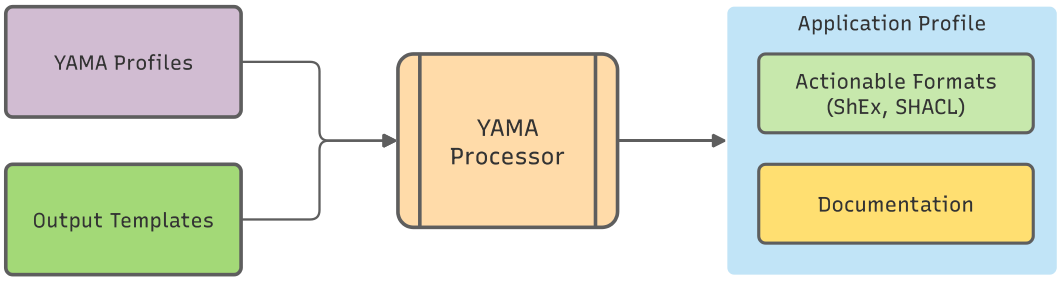


Fig. 5. YAMA processing system.

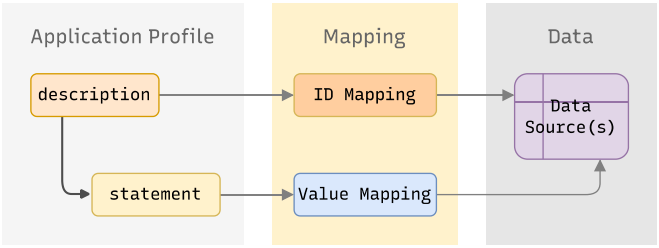


Fig. 6. Application profile to data mapping (Thalhath et al., 2022).

4.6. Challenges and solutions

- RDF Alignment Issues: Some datasets resisted clean alignment with RDF owing to deeply nested structures. On a profile modeling, some best practice recommendations from domain experts were adapted. Reshaping scripts were iteratively developed to clean, and align these datasets.
- Scalability Concerns: As the size and diversity of datasets grew, scalability became a concern. We implemented a streaming generator with Deno to prevent memory allocation issues while working with large datasets.
- Link reconciliation challenges: The 5-star data principle, adhering to linked open data principles, mandates that entities are represented as links. However, aligning all sample datasets to this standard necessitates extensive reconciliation work. To expedite the testing process, the majority of the datasets were utilized without reconciling the entity links, a decision made to balance thoroughness with efficiency in the testing phase.

In conclusion, this methodology balances the strengths of the Dublin Core DSP with the flexibility offered by. Also, this helps to form a base application profile framework and extend to have data mapping capabilities. It aims to contribute significantly to the advancement of

Table 3
Namespaces for the application profile RDF.

Vocabulary	prefix	namespace
RDF Schema	`rdfs:`	http://www.w3.org/2000/01/rdf-schema#
XML Schema	`xsd:`	http://www.w3.org/2001/XMLSchema#
OWL	`owl:`	http://www.w3.org/2002/07/owl#
OWL-DSP	`dsp:`	http://purl.org/metainfo/terms/dsp#
Metabridge Registry	`reg:`	http://purl.org/metainfo/terms/registry#
SHACL	`shacl:`	https://www.w3.org/ns/shacl#

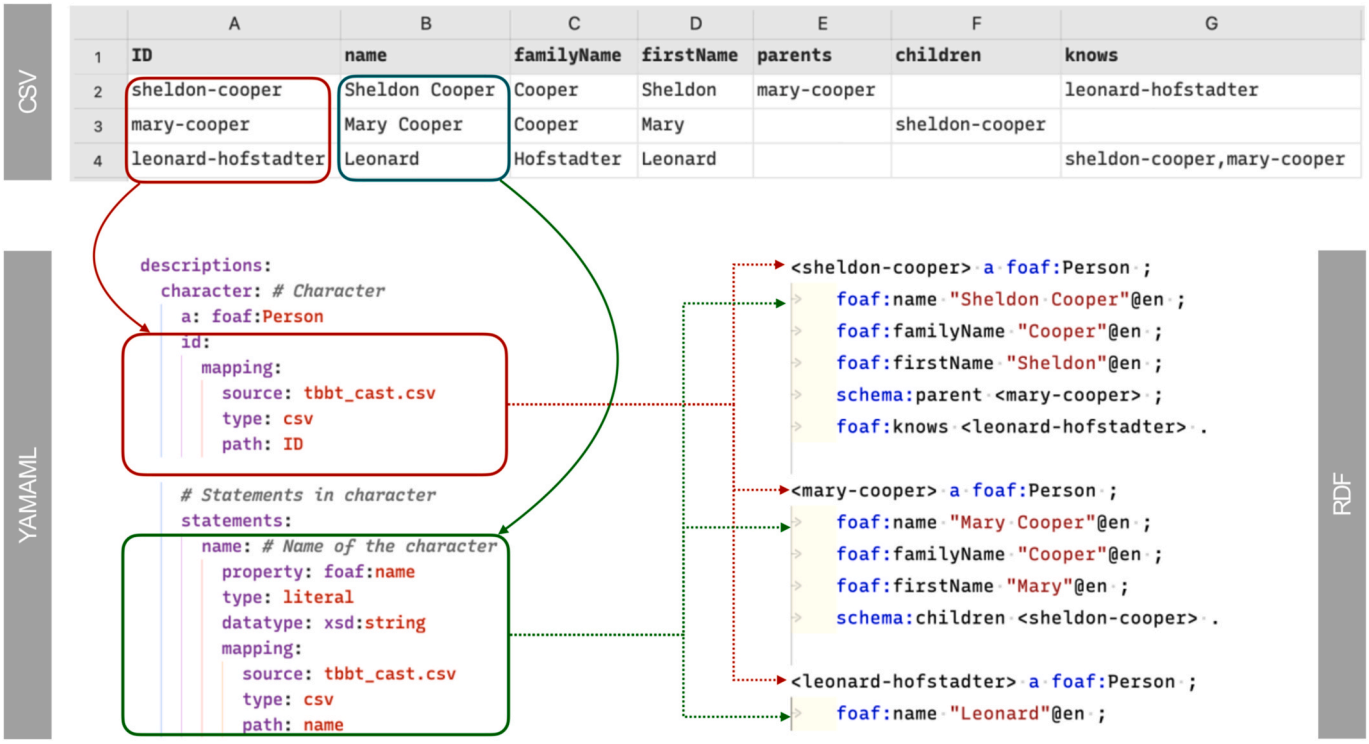


Fig. 7. YAMAML Mapping for the application profile to non-RDF data and the generated RDF.

Table 4
Proposed terms in an RDF application profile.

Section	Element	Term	SHACL
descriptions			
	a		
	description	dsp:DescriptionTemplate	sh:class
	label	rdfs:label	sh:name
	note	rdfs:comment	sh:description
	id	reg:idField	
	type	dsp:valueURIOccurrence	
	statementOrder	reg:statementOrder	
	default		sh:default
	closed		sh:closed
statements			
	statement	dsp:StatementTemplate	sh:property
	label	rdfs:label	sh:name
	note	rdfs:label	sh:description
	property	owl:onProperty	sh:path
	min	owl:minQualifiedCardinality	sh:minCount
	max	owl:maxQualifiedCardinality	sh:maxCount
	type	owl:onDataRange	sh:nodeKind
	datatype	rdf:datatype	sh:datatype
value constraints			
	pattern	xsd:pattern	sh:pattern
	minInclusive	xsd:minInclusive	sh:minInclusive
	maxInclusive	xsd:maxInclusive	sh:maxInclusive
	length	xsd:length	sh:length
	minLength	xsd:minLength	sh:minLength
	maxLength	xsd:maxLength	sh:maxLength
	list	xsd:enumeration	

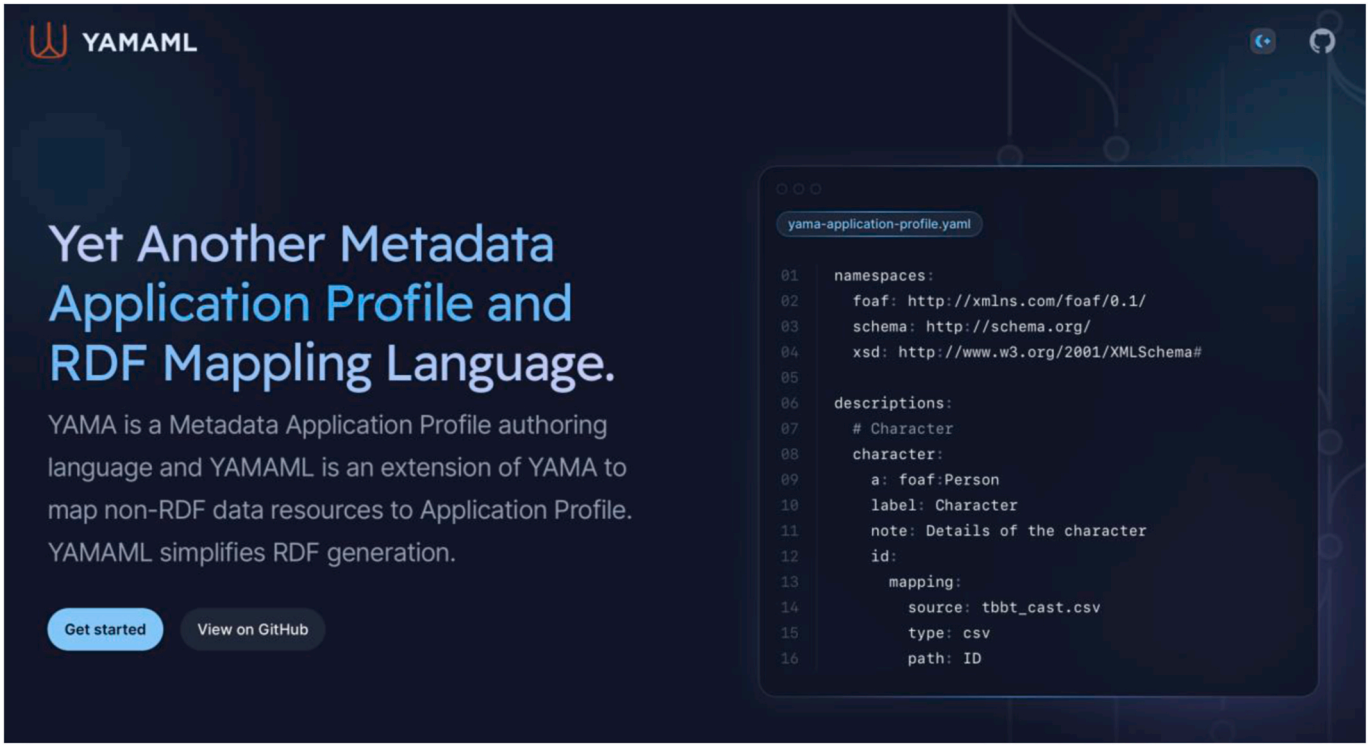


Fig. 8. Yamaml. org website for YAMAML documentation and tools.

semantic interoperability using application profiles in open data publishing and ensures that the data is highly reusable as well as maintains the semantics.

5. Results

5.1. Yet Another Metadata Application Profile (YAMA)

YAMA is a pre-processor that streamlines the process of creating, managing, and publishing Metadata Application Profiles (Thalhath et al., 2019a). It adapts the DublinCore DSP and is influenced by the

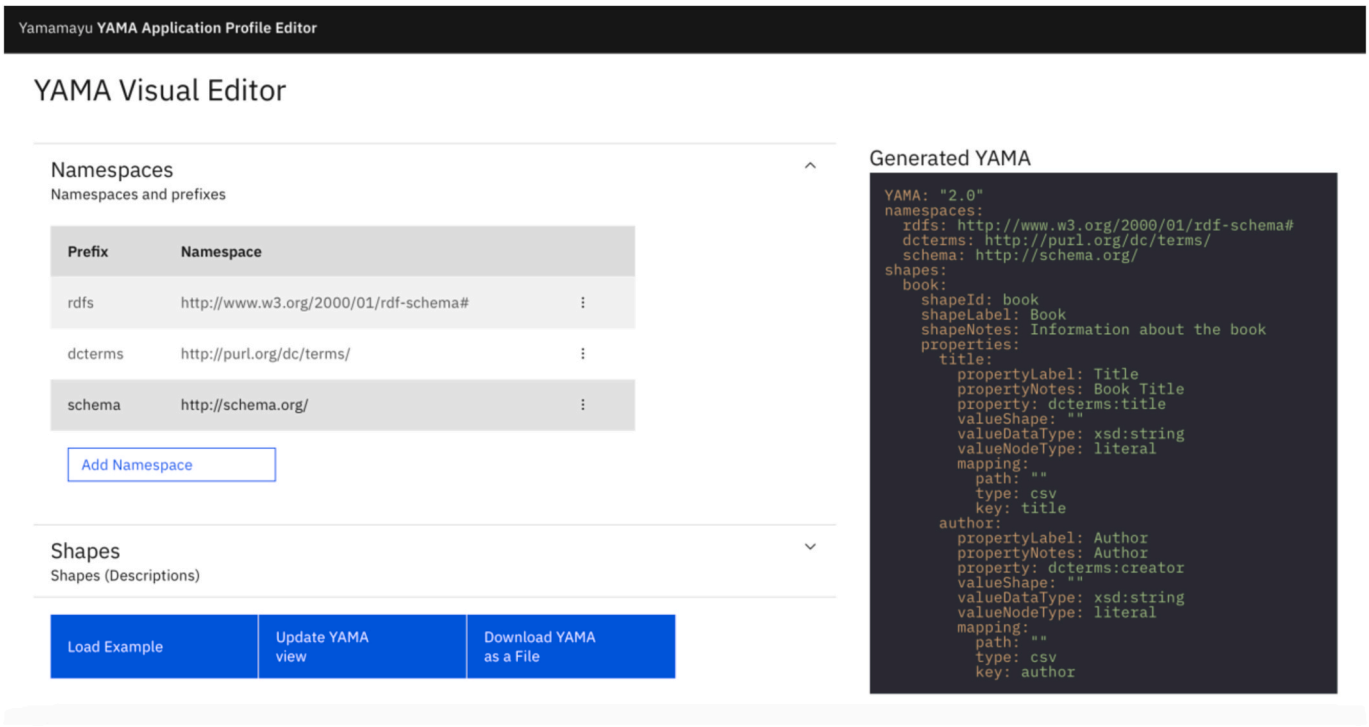


Fig. 9. YAMAML interactive GUI editor webapp

Table 5
Comparison of intermediary formats.

Format	Type	Based on	Description
Simple-DSP	Authoring format	Tabular Values	Generates OWL-DSP
YAMA	Authoring Format	YAML	Intended to generate various standard specifications and generate RDF
LinkML	Data Modeling Format	YAML	Intended to generate various data/ validation formats

Table 6
Comparison of related works with different aspects of this research.

Format	Based on	Expresses in	RDF Generation	Profile	Observation
ShEx	RDF	ShExC or ShExJ (JSON)	No	No	An efficient RDF validation language but lacks other aspects of Application profiles.
DC TAP	RDF, DSP	CSV, TSV	No	Yes	A complete application profile approach but need to convert to validation languages to make actionable.
LinkML	RDF	YAML	Yes	Yes	Not indented to be an application profile, but suitable to be adapted.
YAMA	RDF, DSP	YAML	Yes	Yes	General purpose Application Profile format.
TARQL	RDF	SPARQL	Yes	No	

Simple-DSP format from the MetaBridge project (Nagamori et al., 2011). YAMA is designed to facilitate the creation of profiles using a variety of standards and formats, but it is not intended to be a new standard in itself. Instead, it simplifies the process of profile creation, democratising it for those who may not have in-depth knowledge of metadata application profiles. It is built upon the YAML 1.2 specification syntax and can be processed by any compatible YAML 1.2 parser, although the output might vary based on the specific implementation. With the growing trend of GitHub-based workflows, YAMA seeks to enhance the application profile authoring process. It provides various output formats and seamlessly integrates with frameworks like ShEx, DCAT, and PROV. Serving as an intermediary format, YAMA is equipped to either produce new standard application profile formats or convert existing ones. An overview of the YAMA processing system is illustrated in Fig. 5.

5.2. *YAMAML: an application profile based lightweight RDF mapping language*

The YAMA Mapping Language (YAMAML) has been introduced to enhance the publication of 5-star level open data. The underlying idea is that by employing a profile-driven RDF generation, the process can be optimised, allowing the mapping of multiple non-RDF sources to an RDF application profile with diverse complexities. The application profile for data mapping is illustrated in Fig. 6.

YAMAML subset can be described as descriptions and statements, and the basic parameters are description, `a/rdf:type` (optional), ID mapping, statements statement property, type, datatype, description (if this statement points to another description), and mapping (Value Mapping). An overview of the application profile mapping and generated RDF is demonstrated in Fig. 7.

5.3. *Publishing application profiles in RDF*

The authors formed an application profile to express metadata application profiles by adapting application profile concepts from the existing efforts. The elements required to express the profiles are taken from YAMA and DC-TAP. RDF vocabulary to express the profile is taken from OWL-DSP, Metabridge RDF namespace, RDFS, and OWL. Terms from the SHACL namespace are taken for additional mapping and enforcing constraints. An application profile for application profiles is

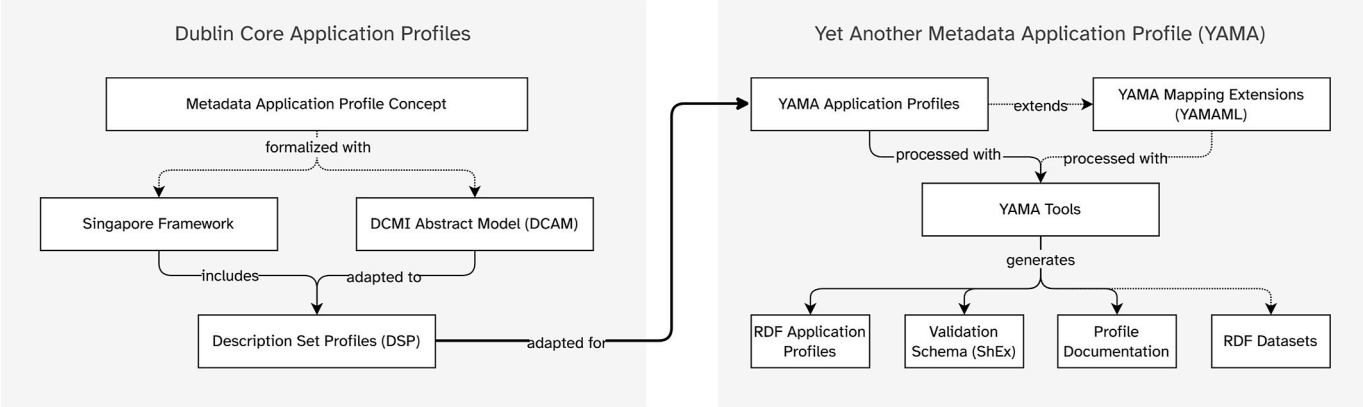


Fig. 10. The relationship between DSP and the results of this study.

derived from the namespaces listed in Table 3.

In this part of the research, our primary objectives encompass several facets of application profile development in RDF. Firstly, we aim to identify a comprehensive set of terms from existing vocabularies that can fully articulate an application profile in RDF. Secondly, we aim to integrate a data validation vocabulary into the application profile, thereby streamlining the generation of validation schemas. Furthermore, it is essential to utilise existing vocabularies in expressing application profiles in RDF to avoid redundancy and promote the reuse of established terms. Lastly, to ensure the interoperability of profiles, we strive to establish a mapping between the existing vocabularies. A comprehensive mapping detailing these relationships is presented in Table 4.

5.4. Implementations

YAMAML (Yet Another Metadata Application Profile and RDF Mapping Language), an emerging language for profiling and generating RDF data, offers a basic tooling suite and comprehensive documentation accessible through the official website https://yamaml.org (Fig. 8). The open-source command line interface (CLI) toolkit enables the generation of relatively large and complex RDF structures through the use of application profiles. A proof-of-concept interactive graphical user interface (GUI) editor for YAMAML is also available as a web application (Fig. 9). The GUI editor provides an alternative to the CLI for users less familiar with command-line interfaces and can be utilized to create and edit YAMAML documents in an intuitive and user-friendly manner.

5.5. Comparison with state of the art

The results of this research have been juxtaposed with the state of the art and pertinent works to gain a comprehensive overview of the findings. YAMA, used here as an application profile intermediary format, is analyzed alongside other similar application profile authoring formats, as detailed in Table 4. Additionally, diverse facets of this study are contrasted with corresponding works, with a concise summary presented in Table 5.

6. Discussion

The development of YAMA is a direct adaptation of DC-DSP and takes inspiration from the Simple-DSP format used in the MetaBridge project (see Table 6). It aims to make application profile concepts accessible to a wider, non-technical audience. While its simplicity restricts its ability to handle complex profiles or use cases, skilled users can still create them manually or through programming. The modular design is planned to evolve into a more object-oriented structure while retaining its simplicity. However, editing YAMA requires a certain level of expertise in using text editors and editing YAML, which may present a challenge for some users.

DCAP and DSP heavily rely on the RDF model and do not cover the profiling of non-RDF data well. There are other attempts to profile data, but this research limits its scope to DSP-based profiles only. As an adaptation of DSP, YAMA also inherits all DSP shortcomings. The RDF application profile mapping is limited to SHACL only, while Shape Expression (ShEX) could be another viable validation option.

The YAMA format is built with extensibility in mind, catering to a range of additional use cases and specific demands (Thalhath et al., 2019b). Given its adaptability, it is poised to be a pivotal format for various RDF and linked open data endeavours. We recognise the potential benefits of integrating an RDF generation mapping language, envisioning it further to propel YAMA's evolution within the application profile domain. While YAMAML has been crafted to handle typical scenarios, it might fall short in addressing intricate tasks demanding advanced data processing and modifications. It's also worth noting that YAMAML isn't devised for reconciling linked data entities. Although it can map linked data using IRI stems, it lacks built-in reconciliation processes. For those seeking reconciliation capabilities and RDF generation in a graphical interface, tools such as OpenRefine are better suited.

YAMA and YAMAML are adaptations of DSP to make application profile authoring and profile-based RDF generation easier. YAMA tools are developed as proof-of-concept tooling for these proposals. More efficient software can be tailored to address various use-case rather than directly adapting YAMA tools. The relationship between DSP and the result of this study is illustrated in Fig. 10.

7. Conclusions

Application Profiles are vital in promoting semantic interoperability among different (meta)data models and harmonizing (meta)data practices across various communities. While numerous tools and mapping languages for working with RDF are available, many have steep learning curves and can be complex for basic purposes. User-friendly tools like OpenRefine, however, enable even novice users to convert their data into RDF format effortlessly. The broad availability of tools catering to diverse user needs and preferences is expected to drive the adoption of RDF and foster open data sharing on the Semantic Web. Although various efforts have been made to create metadata application profile specifications, the practical utilization of RDF application profiles still needs to be improved. Adopting interoperable RDF application profiles can enhance the Findable, Accessible, Interoperable, and Reusable (FAIR) attributes of open datasets, making them easier to discover, use, and reuse. The increasing adoption of Application Profiles within the open data publishing and semantic web communities will improve semantic interoperability and encourage standardized approaches to data description. In summary, this research translates theoretical concepts in

FAIR and open data publishing into practical applications, offering a framework that enhances semantic interoperability. It contributes significantly to the field by developing semantic-aware profiles and tools for open data, thereby improving data openness, usability, and quality. The framework's adaptability and user-friendly design ensure its broad applicability across diverse domains, magnifying its influence in open and FAIR data publishing. As a next step in this research, it is planned to provide comprehensive documentation, examples, more easily adaptable tools, and workflows for semantic-rich open data publishing.

Declaration of generative AI and AI-assisted technologies in the writing process

In preparing this work, the authors utilized Grammarly and the OpenAI API to enhance readability, correct grammar, and refine language use. After employing these services, the authors thoroughly reviewed and edited the content as necessary and took full responsibility for the manuscript's content.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This work was supported by JSPS KAKENHI Grant Number 21K12579.

References

R. Albertoni, D. Browning, S. Cox, A. Gonzalez Beltran, A. Perego, and P. Winstanley, "Data Catalog Vocabulary (DCAT) - Version 2." Accessed: April. 20, 2020. [Online]. Available: https://www.w3.org/TR/vocab-dcat-2/.

Amato, F., Mazzeo, A., Moscato, V., & Picariello, A. (2013). A framework for semantic interoperability over the cloud. In *2013 27th international conference on advanced information networking and applications workshops*. IEEE.

O. Ben-Kiki, C. Evans, and I. döt Net, "YAML Ain't Markup Language (YAML^{TM}) Version 1.2." Accessed: April. 10, 2019. [Online]. Available: https://yaml.org/spec/1.2/spec.html.

Berners-Lee, T. (2006). "Linked data," *W3C des. Issues*. http://www.w3.org/DesignIssues/LinkedData.html.

Bestek, M., Grönvall, E., & Saad-Sulonen, J. (2022). Commoning semantic interoperability in healthcare. *International Journal of the Commons, 16*(1), 225–242.

Bezuidenhout, L. (2020). Being fair about the design of FAIR data standards. *Digit. Gov. Res. Pract., 1*(3), 1–7.

Bizer, C., Heath, T., & Berners-Lee, T. (2011). Linked data: The story so far. In *Semantic services, interoperability and web applications: Emerging concepts* (pp. 205–227). IGI global.

Car, N. (2019). *"The profiles vocabulary," W3C*. W3C Note.

Cathy, S. (2018). Lin and hsin-chang yang, "evaluating semantic interoperability of government open data portals,". *Jt. Int. Conf. Semantic Technol.*

Coyle, K., Baker, T., Barker, P., Huck, J., Riesenberg, B., & Thalhath, N. (2023). *DC tabular application profiles (DC TAP)*. Dublin Core Metadata Initiative [Online]. Available: https://www.dublincore.org/specifications/dctap/primer/.

Cyganiak, R. (2015). *Tarql (sparql for tables): Turn csv into rdf using sparql syntax* [Online]. Available: http://tarql.github.io.

Das, S., Sundara, S., & Cyganiak, R. (2012). *"R2RML: RDB to RDF mapping language," W3C, W3C recommendation*.

Davies, J., Harris, S., Crichton, C., Shukla, A., & Gibbons, J. (2008). Metadata standards for semantic interoperability in electronic government. In *Proceedings of the 2nd international conference on Theory and practice of electronic governance*. ACM.

DDI Lifecycle 3.3, DDI Lifecycle 3.3 | Data Documentation Initiative. Accessed: May 1, 2023. [Online]. Available: https://ddialliance.org/Specification/DDI-Lifecycle/3.3/.

Dimou, A., Vander Sande, M., Colpaert, P., Verborgh, R., Mannens, E., & Van de Walle, R. (2014). Rml: A generic language for integrated rdf mappings of heterogeneous data. *Ldow, 1184*.

F. Enoksson, "DCMI: A MoinMoin Wiki Syntax for Description Set Profiles." Accessed: March. 11, 2019. [Online]. Available: http://www.dublincore.org/specifications/dublin-core/dsp-wiki-syntax/.

Gal, A. (1999). Semantic interoperability in information services. *ACM SIGMOD Rec, 28*(1), 68–75.

George, G., Haas, M. R., & Pentland, A. (2014). Big data and management. *Academy of Management Journal, 57*(2), 321–326. https://doi.org/10.5465/amj.2014.4002

GitHub - AtesComp/rdf-transform: RDF Transform is an extension for OpenRefine to transform data into RDF formats. — github.com." [Online]. Available: https://github.com/AtesComp/rdf-transform.

Harvey, F., Kuhn, W., Pundt, H., Bishr, Y., & Riedemann, C. (1999). Semantic interoperability: A central issue for sharing geographic information. *The Annals of Regional Science, 33*(2), 213–232.

Hasnain, A., & Rebholz-Schuhmann, D. (2018). Assessing FAIR data principles against the 5-star open data principles. In A. Gangemi, A. L. Gentile, A. G. Nuzzolese, S. Rudolph, M. Maleshkova, H. Paulheim, J. Z. Pan, & M. Alam (Eds.), *Lecture notes in computer scienceThe semantic web: ESWC 2018 satellite events* (pp. 469–477). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-98192-5_60.

Heery, R., & Patel, M. (2000). Application profiles: Mixing and matching metadata schemas. *Ariadne, 25* [Online]. Available: http://www.ariadne.ac.uk/issue/25/app-profiles/. (Accessed 18 March 2018).

Heyvaert, P., De Meester, B., Dimou, A., & Verborgh, R. (2018). Declarative rules for linked data generation at your fingertips. In A. Gangemi, A. L. Gentile, A. G. Nuzzolese, S. Rudolph, M. Maleshkova, H. Paulheim, J. Z. Pan, & M. Alam (Eds.), *The semantic web: ESWC 2018 satellite events* (Vol. 11155, pp. 213–217). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-98192-5_40.

Hoxha, J., & Brahaj, A. (2011). Open government data on the web: A semantic approach. In *2011 international conference on emerging intelligent data and web technologies*. IEEE.

Hyvönen, E., Tuominen, J., Alonen, M., & Mäkelä, E. (2014). Linked data Finland: A 7-star model and platform for publishing and Re-using linked datasets. In V. Presutti, E. Blomqvist, R. Troncy, H. Sack, I. Papadakis, & A. Tordai (Eds.), *Lecture notes in computer science: Vol. 8798. The semantic web: ESWC 2014 satellite events* (pp. 226–230). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-11955-7_24, 8798.

ITU-T. (2005). *Definition of Open Standards*. ITU. https://www.itu.int/en/ITU-T/ipr/Pages/open.aspx.

Kontokostas, D., & Knublauch, H. (2017). *"Shapes constraint language (SHACL)," W3C, W3C recommendation*.

Meredith, J., Whitehead, N., & Dacey, M. (2022). Aligning semantic interoperability frameworks with the FOXS stack for FAIR health data. *Methods of Information in Medicine, 62*(1), e39–e46.

Moxon, S. A., et al. (2021). *The linked data modeling language (LinkML): A general-purpose data modeling framework grounded in machine-readable semantics*. ICBO.

Nagamori, M., Kanzaki, M., Torigoshi, N., & Sugimoto, S. (2011). *Meta-bridge: A development of metadata information infrastructure in Japan* (Vol. 6).

M. Nilsson, "DCMI: Description Set Profiles: A constraint language for Dublin Core Application Profiles." Accessed: March. 1, 2019. [Online]. Available: http://www.dublincore.org/specifications/dublin-core/dc-dsp/.

M. Nilsson, T. Baker, and P. Johnston, "DCMI: The Singapore Framework for Dublin Core Application Profiles." Accessed: May 18, 2019. [Online]. Available: http://dublincore.org/specifications/dublin-core/singapore-framework/.

OECD. (2021). *Data portability, interoperability and digital platform competition*. OECD Compet. Comm. Discuss. Pap. [Online]. Available: http://oe.cd/dpic. (Accessed 3 October 2022).

OpenAPI Initiative, "OpenAPI Specification." Accessed: May 15, 2019. [Online]. Available: https://swagger.io/specification/.

Osumi-Sutherland, D., Courtot, M., Balhoff, J. P., & Mungall, C. (2017). Dead simple OWL design patterns. *Journal of Biomedical Semantics, 8*(1), 18. https://doi.org/10.1186/s13326-017-0126-0

Paul, M., & Ghosh, S. (2012). *A framework for semantic interoperability for distributed geospatial repositories*. Comput. Inform.

Polleres, A., & Steyskal, S. (2015). Semantic web standards for publishing and integrating open data. In *Standards and standardization* (pp. 1–20). IGI Global.

Raskin, R., Pan, M. J., & Mattmann, C. (2004). *Semantic interoperability for earth science data*.

Rowley, J. (2007). The wisdom hierarchy: Representations of the DIKW hierarchy. *Journal of Information Science, 33*(2), 163–180. https://doi.org/10.1177/0165551506070706

Schriml, L. M., Lichenstein, R., Bisordi, K., Bearer, C., Baron, J. A., & Greene, C. (2023). Modeling the enigma of complex disease etiology. *Journal of Translational Medicine, 21*(1), 148. https://doi.org/10.1186/s12967-023-03987-x

Sciore, E., Siegel, M., & Rosenthal, A. (1994). Using semantic values to facilitate interoperability among heterogeneous information systems. *ACM Transactions on Database Systems, 19*(2), 254–290.

Seaborne, A., & Harris, S. (2013). *"SPARQL 1.1 query language," W3C, W3C recommendation*.

Soiland-Reyes, S., Castro, L. J., Garijo, D., Portier, M., Goble, C., & Groth, P. (2022). Updating linked data practices for FAIR digital object principles. *Res. Ideas Outcomes, 8*(Oct).

Strawn, G. (2019). Open science, business analytics, and FAIR digital objects. In *2019 IEEE 43rd annual computer software and applications conference (COMPSAC)*. IEEE.

Stupnikov, S., & Kalinichenko, L. (2019). Extensible unifying data model design for data integration in FAIR data infrastructures. In *Communications in computer and information science* (pp. 17–36). Springer International Publishing.

Tandy, J., Herman, I., & Kellogg, G. (2015). *"Generating RDF from tabular data on the web," W3C, W3C recommendation*.

Thalhath, N., Nagamori, M., & Sakaguchi, T. (2020). MetaProfiles - a mechanism to express metadata schema, privacy, rights and provenance for data interoperability. In E. Ishita, N. L. S. Pang, & L. Zhou (Eds.), *Lecture notes in computer scienceDigital libraries at times of massive societal transition* (pp. 364–370). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-64452-9_34.

Thalhath, N., Nagamori, M., & Sakaguchi, T. (2022). Yamaml: An application profile based lightweight RDF Mapping Language. In Y.-H. Tseng, M. Katsurai, & H. N. Nguyen (Eds.), *Lecture notes in computer scienceFrom born-physical to born-virtual: Augmenting intelligence in digital libraries* (pp. 412–420). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-031-21756-2_32.

Thalhath, N., Nagamori, M., Sakaguchi, T., & Sugimoto, S. (2019a). Yet another metadata application profile (YAMA): Authoring, versioning and publishing of application profiles. In *Proceedings of the 2019 international conference on Dublin core and metadata applications* (pp. 114–125). DCMI'19. Seoul, South Korea: Dublin Core Metadata Initiative.

Thalhath, N., Nagamori, M., Sakaguchi, T., & Sugimoto, S. (2019b). Authoring formats and their extensibility for application profiles. In A. Jatowt, A. Maeda, & S. Y. Syn (Eds.), *Lecture notes in computer scienceDigital libraries at the crossroads of digital information for the future* (pp. 116–122). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-34058-2_12.

Thalhath, N., Nagamori, M., Sakaguchi, T., & Sugimoto, S. (2020). Metadata application profile provenance with extensible authoring format and PAV ontology. In X. Wang, F. A. Lisi, G. Xiao, & E. Botoeva (Eds.), *Lecture notes in computer scienceSemantic technology* (pp. 353–368). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-41407-8_23.

Thalhath, N., Nagamori, M., Sakaguchi, T., & Sugimoto, S. (2021). Wikidata centric vocabularies and URIs for linking data in semantic web driven digital curation. In E. Garoufallou, & M.-A. Ovalle-Perandones (Eds.), *Communications in computer and information scienceMetadata and semantic research* (pp. 336–344). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-71903-6_31.

Thornton, K., et al. (2019). Using shape expressions (ShEx) to share RDF data models and to guide curation with rigorous validation. In P. Hitzler, M. Fernández, K. Janowicz, A. Zaveri, A. J. G. Gray, V. Lopez, A. Haller, & K. Hammar (Eds.), *Lecture notes in computer scienceThe semantic web* (pp. 606–620). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-21348-0_39.

Van Assche, D., Delva, T., Heyvaert, P., De Meester, B., & Dimou, A. (2021). Towards a more human-friendly knowledge graph generation & publication. In *International semantic web conference (ISWC) 2021: Posters, demos, and industry tracks*.

Vogt, L., & Extending FAIR to FAIREr. (2023). *Cognitive interoperability and the human explorability of data and metadata*. ArXiv.

The ODC, International Open Data Charter. [Online]. Available: https://opendatacharter.net/principles/.

P. Walsh and R. Pollock, "Data Package." Accessed: May 23, 2020. [Online]. Available: https://specs.frictionlessdata.io/data-package/#language.

Wilkinson, M. D., et al. (2017). Interoperability and FAIRness through a novel combination of Web technologies. *PeerJ Comput. Sci., 3*, e110.

Wilkinson, M. D., et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data, 3*(1). https://doi.org/10.1038/sdata.2016.18. Art. no. 1.

Wood, D., Lanthaler, M., & Cyganiak, R. (2014). *"RDF 1.1 concepts and Abstract syntax," W3C*. W3C Recommendation.