ADVANCES IN KNOWLEDGE ORGANIZATION
Proceedings of International ISKO Conference - ISSN: 0938-5495 - 2026 - São Paulo

Using Schema and Value Vocabularies to Structure Information Applications

Joseph Busch

Taxonomy Strategies, USA

Abstract

Application profiles provide a framework for structuring information related to a particular application but are not considered to be knowledge organization systems in themselves. Nevertheless application profiles contain two kinds of vocabularies—metadata schemas and controlled vocabularies. While the vision of the application profile was to enable interoperability, the practice of documenting how the requirements could be met may be just as important. Application profiles may also be a useful framework for integrating vocabularies into generative AI to improve its reliability and transparency.

1.0 Introduction

An application profile defines a specific set of requirements, settings, and metadata for a particular application. To ensure compatibility and functionality, the profile adapts general standards or frameworks, for example, using Dublin Core Metadata Terms (2020) for bibliographic description and Friend of a Friend (FOAF)[1] for named entity properties rather than specifying a whole schema from scratch. Heery and Patel (2000) provide this frequently cited definition of application profiles as "schemas which consist of data elements drawn from one or more namespaces, combined together by implementors, and optimised for a particular local application." They explain that the purpose of a namespace is to:

1) Identify the authority for the schema,
2) Provide unique (and persistent) identifiers for the data elements, and
3) Define the particular data elements.

Heery and Patel refer to the data element set as "a vocabulary". They also explain that a distinguishing characteristic of an application profile is that it may specify particular controlled vocabularies for use in populating specific data elements. Thus in this context, an application profile itself is not a knowledge organization system (KOS), but the data element set is.

The SEMIC Support Center provides a more data-centric definition that refers to application profiles as "semantic data specifications" referring to the semantic web. The goal of the

[1] FOAF (Friend of a Friend) is an RDF vocabulary of terms related to the description of people. (Brickley and Miller 2014)

semantic web was to make Internet data machine-readable by encoding semantics with Resource Description Framework (RDF)[2] and Web Ontology Language (OWL).[3]

"An Application Profile (AP) is a data specification aimed to facilitate the data exchange in a well-defined application context. It re-uses concepts from one or more semantic data specifications, while adding more specificity, by identifying mandatory, recommended, and optional elements, addressing application needs, and providing recommendations for controlled vocabularies to be used." (SEMIC Support Centre 2023).

The World Wide Web Consortium (W3C) Data Exchange Working Group developed PROF "The Profiles Vocabulary" (2019)[4] and "Guidance on Publishing Application Profiles of Vocabularies" (2025). A W3C application profile defines a particular way of using one or more base specifications to achieve a specific function or interoperability goal within a community of practice. An application profile doesn't create entirely new web standards but rather specifies how existing W3C Recommendations (like RDF, SKOS[5], or DCAT[6] specifications) should be used in a specific context. It specifies the components needed to define the profile and also how to validate conformance with the profile. This might involve limiting the use of certain features, requiring specific data types or vocabularies, or defining how different specifications interact.

While each of these definitions is concerned with facilitating the interoperability of applications through transparent description, some emphasize data over more general content, and all of them emphasize the container more than the content. All of these definitions refer to the schema as a vocabulary and in some cases make reference to controlled vocabularies for use in a metadata element. While they refer to profiles as semantic schemes they are not generally concerned with the meaning of content, but whether the container or data element structurally conforms to the specification.

2.0 Schema and Value Vocabularies

The specifications refer to profiles as semantic schemes although application profiles are not generally considered to be knowledge organization systems in themselves. But application profiles contain two kinds of vocabularies—metadata schemas and controlled vocabularies both of which are considered KOS.[7] Metadata schemas define the elements and rules for metadata, while controlled vocabularies, such as taxonomies, thesauri, and

[2] Resource Description Format (RDF) is a method to structure data as a graph. An RDF graph is structured as a "triple" consisting of two nodes – a subject and object – connected by a predicate (sometimes called an "edge"). Each component of a triple can have a unique persistent identifier or simply be a label (called a "literal").

[3] Web Ontology Language (OWL) is a method to structure and represent things, groups of things, and relations between things.

[4] "The Profiles Vocabulary" (2019) is an RDF vocabulary created to allow the machine-readable description of profiles of specifications for information resources.

[5] "SKOS Simple Knowledge Organization System" (2009) is a widely used common data model for sharing and linking knowledge organization systems on the Web.

[6] "DCAT Data Catalog Vocabulary" (2024) is an RDF vocabulary designed to facilitate interoperability between data catalogs published on the Web.

ontologies provide a consistent set of terms and sometimes relationships to describe what the information is about in a consistent way.

The purpose of metadata schemas is to define the types and structures of metadata used for describing a collection of resources. For example, Dublin Core[8] is a well-known metadata schema that provides a standard set of Metadata elements for describing digital resources. Metadata schemas are prescriptive, outlining how information should be organized and presented to ensure interoperability.[9]

The purpose of controlled vocabulary is to provide a controlled set of terms and relationships for describing the content itself, ensuring consistency and clarity. For example,

- **Taxonomies** organize concepts into hierarchical "is a" or "part of" relationships, creating a structured tree for concepts.
- **Thesauri** are a type of controlled vocabulary that includes preferred terms, synonyms (variant terms), broader terms, narrower terms, and related terms and comply with ISO 25964 Part 1 (ISO 2011).
- **Authority Files** are controlled vocabularies specifically used for authoritative lists of names of people, organizations, places, and other entities.
- **Ontologies** are more complex semantic schemes that define concepts and their relationships with greater specificity than a thesaurus or taxonomy.

Controlled vocabularies aim to unify meaning and improve information retrieval by ensuring that terms are used consistently and unambiguously across a collection.

3.0 Constructing Application Profiles

Based on the “Singapore Framework” shown in Figure 1, Coyle and Baker’s (2009) Guidelines for Dublin Core Application Profiles (DCAP) are a primer that describes the types of documents that an application profile should contain including:

- Functional Requirements that define the actions and features that are needed to fulfill the use cases,
- Domain Model identifies the key entities and relationships between them that are involved in the application,
- Description Set that enumerates the data model,
- Usage Guidelines that provide rules for using the data, and

[7] See comprehensive discussion of knowledge organization systems in *ISKO Encyclopedia of Knowledge Organization* (Mazzocchi 2018).
[8] The Dublin Core metadata schema is generally regarded as the 15 properties in the /elements/1.1/ namespace. https://www.dublincore.org/specifications/dublin-core/dcmi-terms/#section-3.
[9] See comprehensive discussion of interoperability in *ISKO Encyclopedia of Knowledge Organization* (Zeng 2019).

- Syntax Guidelines that describe how the data needs to be encoded for data processing.

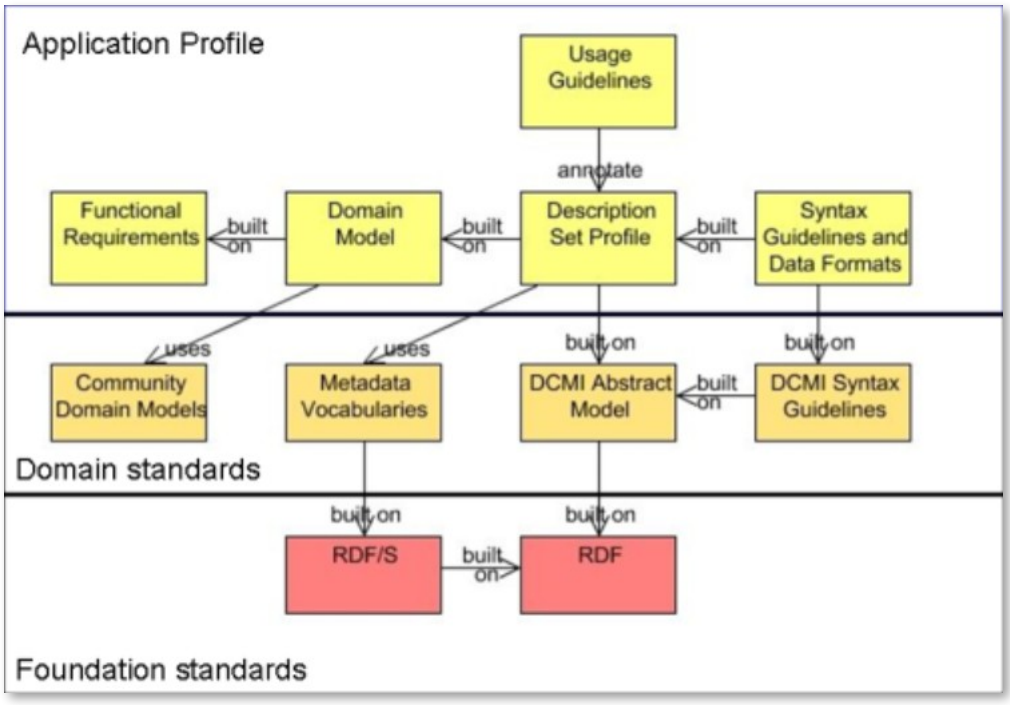

Figure 1. The Singapore Framework for Dublin Core Application Profiles[10]

A useful paper by Malta and Baptista (2013) describes a specific method for the development of DCAP based on the Design Science Research (DSR) methodological approach. DSR is a research paradigm applied to the development and validation of prescriptive knowledge in information science.[11] This practical application of DCAP provides further insight into the process and identifies, for example, additional inputs and methods for validation of specific application profile documents.

4.0 The Purpose of Application Profiles

The primary goal of an application profile is to facilitate out-of-the-box interoperability between different systems or applications within a defined domain. By adhering to a common profile, developers can ensure that their implementations can seamlessly exchange information and interact with each other.

[10] The Singapore Framework for Dublin Core Application Profiles is a framework for designing metadata applications for maximum interoperability and for documenting such applications for maximum reusability. (Nilsson et al. 2008)

[11] Design science (methodology). Wikipedia. https://en.wikipedia.org/wiki/Design_science_(methodology).

Application profiles are often developed to address the specific needs of a particular community or industry sector. For example, a profile might be created for sharing learning resource metadata (LRMI specification)[12], for creating Dublin Core applications (DCAP)[13], for tabular Dublin Core application profiles (DC TAP)[14], or for sharing knowledge organization systems (NKOS AP)[15].

Nevile and Méndez (2015, 76) explore the question "Why do you want an application profile?" This paper suggests that the inherent nature of RDF and proliferation of RDF tools such as triplestore databases[16] make an application profile unnecessary for semantic interoperability.[17] While general semantic interoperability does not require an application profile, the value of the application profile is to establish good practices for developing semantic applications. The purpose of an application profile is to document a set of requirements and specifications for how those requirements are to be met. While the vision of the application profile was to enable interoperability, the practice of documenting how the requirements could be met may be just as important.

This is particularly important in the context of current artificial intelligence (AI) and data management practices for a number of reasons:

1) In the current large language model (LLM) and generative AI paradigm "everything is data". This means that unstructured content such as untagged text and rich media as well as structured content and data itself are all within processing scope.
2) Generative AI does not truly understand language (or see images or hear music). It is a complicated statistical model that generates a response to a prompt (what you submit) token-by-token, each token predicting the next. Even though AI responses appear to be compelling, the reality is that the same prompt returns a different response each time. The results are difficult to refine because the mechanisms by which they were generated are not apparent.
3) The AI user interface is natural language which lowers the bar to use but makes it more difficult to define what is a good practice than with a structured language (like Java, SQL, etc.).

5.0 Improving the Reliability of AI Using Value and Schema Vocabularies

[12] LRMI Learning Resource Metadata Innovation specification (2014) is a collection of RDF classes, properties and concept schemes for markup and description of educational resources.
[13] See *Dublin Core Application Profile Guidelines*. (Coyle and Baker 2009)
[14] DCTAP Dublin Core Tabular Application Profiles (n.d.) is a simple model that can be used to define application profiles.
[15] Networked Knowledge Organization Systems Dublin Core Application Profile NKOS AP (2019) is a set of RDF classes and properties to describe knowledge organization systems.
[16] A triplestore is a database used for the storage and retrieval of triples using semantic queries.
[17] Nevile and Méndez (2015) do not ask whether the semantic web is necessary, instead they question the necessity of application profiles in an RDF data world. Instead, they should have asked whether the vision for the semantic web needed to be revised. The vision of machine-understandable public data on a global scale was primary to the semantic web. The reality is that the persistence of and access to public data on a global scale is not practically sustainable.

To improve the reliability of search you need to use better search terms. One way to do this is to look for and choose search terms from vocabularies. The same is true for improving the reliability of AI. Figure 2 illustrates a simple approach to integrating vocabularies into AI prompts which is similar to integrating vocabularies in information retrieval queries:

- Incorporate synonyms (and quasi-synonyms) in the prompt using a domain thesaurus;
- Broaden or narrow the concepts in the prompt using a hierarchical taxonomy; and
- Provide specific contexts using metadata schemas such as content type, geography, language, time, process, or other contexts.

While it's possible to manually refer to vocabularies while writing prompts, LLMs integrate with other systems through the "context window". The context window is the input that the LLM processes. LLMs predict the next token based on the input sequence, and by manipulating this input, one can affect the model's output. The ChatGPT context window is capped at 32,000 tokens. A token is roughly equivalent to 4 characters or about three quarters of a word. But processing large context windows can be computationally expensive and slow.

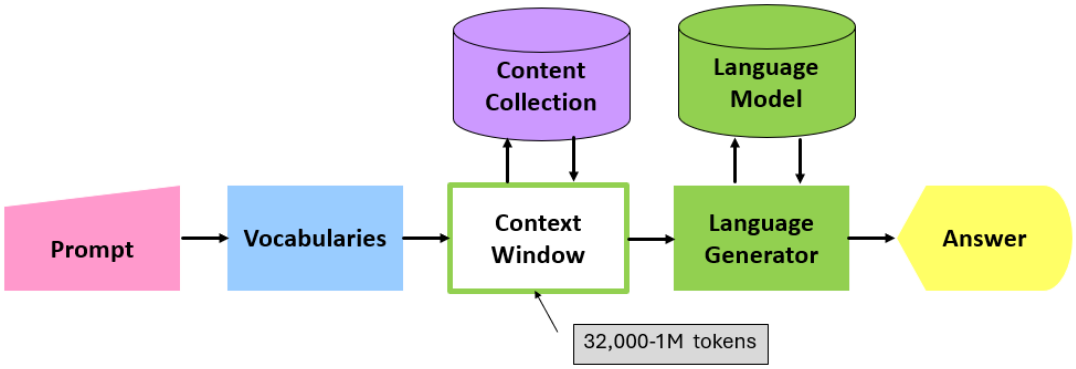

Figure 2. Simple integration of vocabularies with generative AI application.

An application programming interface (API) allows for development of applications that can integrate additional resources such as profile vocabularies and search results into prompts. The ChatGPT context window is up to a million tokens when accessed via the API. In this context, vocabularies can also be used to:

- Guide language generation results by identifying entities in the output, link entities to knowledge graphs, and connect models to knowledge bases;
- Index source content to improve document recall and precision; and
- Annotate output to improve explainability and trustworthiness.

Retrieval augmented generation (RAG) involves searching for relevant facts and inserting them into the context window to improve the model's answers. The approach can be expanded to include various data sources beyond search engines, such as an indexed domain specific collection, databases, and calendars, making it a complex but powerful

tool for enhancing LLM performance. The response may include annotations, such as source citations or bibliographic references from the domain specific collection content. A good way to think about RAG is as a method for pre-processing LLM prompts to make them more responsive to domain specific questions.

A knowledge graph is a representation of the relationships among named entities. A knowledge graph may be queried in the RAG method to improve the information retrieval with equivalent terms but especially to provide associative, typed related concepts. It can also be integrated as a post-process to an LLM response to identify possible refinements through reasoning, decomposition, verification, or other methods.

RAG applications may use information retrieval in different ways. In particular, they may not necessarily index content using vocabularies. They may use vector search to identify similar content statistically without any reference to vocabularies. A better way (to achieve more reliable and measurably better recall and precision) is to use vectors to map content to vocabularies, and then to use vocabularies to identify similar content.

6.0 Summary

This paper discussed how application profiles provide a framework for structuring information related to particular applications. It discussed the important specifications for profile vocabularies developed by the Dublin Core and W3C communities. These specifications refer to profiles as semantic schemes although application profiles are not generally considered to be knowledge organization systems in themselves. But application profiles contain two kinds of vocabularies—metadata schemas and controlled vocabularies both of which are considered KOS. While the vision of the application profile was to enable interoperability, the practice of documenting the requirements and how they can be met may be just as important. Profiles may also be a useful framework for integrating both types of vocabularies into AI applications to improve their reliability and transparency.

Acknowledgement

The author acknowledges the use of Google Gemini to assist with research and to help with an early draft of Section 1.0 Introduction of this paper.

References

Brickley, Dan, and Libby Miller. 2014. “FOAF Vocabulary Specification 0.99: Namespace Document.” January 14, 2014. https://xmlns.com/foaf/spec/

Coyle, Karen, and Thomas Baker. 2009. “Guidelines for Dublin Core Application Profiles.” DCMI Recommended Resource, May 18, 2009. https://www.dublincore.org/specifications/dublin-core/profile-guidelines/

"Data Catalog Vocabulary (DCAT) - Version 3." 2024. W3C Recommendation, August 22, 2024. https://www.w3.org/TR/vocab-dcat-3/

"DCMI Metadata Terms." 2020. DCMI Usage Board. DCMI Recommendation, January 20, 2020. https://www.dublincore.org/specifications/dublin-core/dcmi-terms/

"DCTAP." n.d. Dublin Core Metadata Initiative. https://www.dublincore.org/specifications/dctap/

Heery, Rachel, and Manjula Patel. 2000. "Application Profiles: Mixing and Matching Metadata Schemas." Ariadne, no. 25 (September 24, 2000). https://www.ariadne.ac.uk/issue/25/app-profiles/

International Organization for Standardization. 2011. Information and documentation — Thesauri and interoperability with other vocabularies — Part 1: Thesauri for information retrieval. ISO 25964-1:2011. https://www.iso.org/standard/53657.html

"LRMI Specification Version 1.1." 2014. Association of Educational Publishers and Creative Commons. DCMI Community Specification, October 23, 2014. https://www.dublincore.org/specifications/lrmi/1.1/

Malta, Mariana Curado, and Ana Alice Baptista. 2013. "A Method for the Development of Dublin Core Application Profiles (Me4DCAP V0.1): A Description." In Proceedings of the 2013 International Conference on Dublin Core and Metadata Applications, edited by Muriel Foulonneau and Kai Eckert, 90–103. Lisbon: Dublin Core Metadata Initiative. https://doi.org/10.23106/dcmi.952136137

Mazzocchi, Fulvio. 2018. "Knowledge Organization System (KOS): An Introductory Critical Account." Knowledge Organization 45: 54–78. Also available in ISKO Encyclopedia of Knowledge Organization, edited by Birger Hjørland and Claudio Gnoli. https://www.isko.org/cyclo/kos

Nevile, Liddy, and Eva M. Méndez. 2015. "Do We Need Application Profiles? Reflections and Suggestions from Work in DCMI and ISO/IEC." In Proceedings of the 2015 International Conference on Dublin Core and Metadata Applications, edited by Mariana Curado Malta and Silvana Aparecida Borsetti Gregório Vidotti, 76–86. São Paulo: Dublin Core Metadata Initiative. https://doi.org/10.23106/dcmi.952137032

Nilsson, Mikael, Tom Baker, and Pete Johnston. 2008. "The Singapore Framework for Dublin Core Application Profiles." Dublin Core Metadata Initiative, January 14, 2008. https://www.dublincore.org/specifications/dublin-core/singapore-framework/

SEMIC Support Centre. 2023. "Application Profiles: What Are They and How to Model and Reuse Them Properly? A Look through the DCAT-AP Example." December 2023. https://interoperable-europe.ec.europa.eu/collection/semic-support-centre/application-profiles-what-are-they-and-how-model-and-reuse-them-properly-look-through-dcat-ap

"SKOS Simple Knowledge Organization System Reference." 2009. W3C Recommendation, August 18, 2009. https://www.w3.org/TR/skos-reference/

"The Profiles Vocabulary." 2019. W3C Working Group Note, December 18, 2019. https://www.w3.org/TR/dx-prof/

W3C Dataset Exchange Working Group. 2025. "Guidance on Publishing Application Profiles of Vocabularies." Last edited October 10, 2025. https://www.w3.org/2017/dxwg/wiki/Guidance_on_publishing_application_profiles_of_vocabularies

Zeng, Marcia Lei. 2019. "Interoperability." Knowledge Organization 46: 122–146. Also available in ISKO Encyclopedia of Knowledge Organization, edited by Birger Hjørland and Claudio Gnoli. https://www.isko.org/cyclo/interoperability

Zeng, Marcia L., and Maja Žumer. 2019. "Networked Knowledge Organization Systems Dublin Core Application Profile (NKOS AP)." Version 0.2. Last modified November 3, 2019. DCMI NKOS Task Group. https://nkos.dublincore.org/nkos-ap.html