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

VRA Core as an Integrative Axis in the Data Model Reformulation of ARQUIGRAFIA: A Critical Analysis

Zaira Regina Zafalon

Federal University of São Carlos, Brazil

Cibele Araújo Camargo Marques dos Santos

University of São Paulo, Brazil

Vânia Mara Alves Lima

University of São Paulo, Brazil

Henrique Jungues

FAPESP Scholarship, Brazil

Ana Ritschel Ribeiro

FAPESP Scholarship, Brazil

Abstract

This research aims to investigate the cross-cutting role played by the VRA Core metadata standard in the redesign of the data model of ARQUIGRAFIA, a collaborative platform for images of architecture and urban spaces. The study analyzes the relationship between metadata standards, interoperability, and data conversion in legacy systems; demonstrates the function of VRA Core as a conceptual framework for the analysis of work and image entities, and as a guideline for the critical evaluation of legacy data; and reports how VRA Core provided the framework for the new ARQUIGRAFIA platform, by portraying specific analyses, uses, and adaptations of the standard.

1.0 Introduction

Currently, the proliferation of image platforms stems from the digitization of rare collections and the explosion of born-digital images, driven by the popularization of cell phones, the Internet, and social networks. These environments aim to preserve originals while ensuring web access. Additionally, collaborative platforms allow users to participate in selecting, including, and processing information. In this context, Knowledge Organization (KO) faces increased challenges in establishing principles for identifying and representing image objects, as well as in supporting the development of systems for searching and retrieving these materials. This research investigates the cross-cutting role of the VRA Core metadata standard in redesigning the ARQUIGRAFIA platform's data model. The study [1] analyzes the relationship between metadata standards,

interoperability, and data conversion in legacy systems; [2] demonstrates VRA Core as a conceptual framework for analyzing work and image entities and for critically evaluating legacy data; and [3] reports how VRA Core served as the scope for the new ARQUIGRAFIA, portraying specific analyses, uses, and adaptations of the standard. The methodology combines bibliographic research with comparative analysis of data structures, metadata elements, and representation formats for visual resources, focusing on how these components support semantic consistency in the migration from legacy systems to a structured metadata model.

ARQUIGRAFIA (https://www.arquigrafia.org.br) is a collaborative, open-access, public, non-profit digital environment dedicated to the study of visual culture and architectural images of Brazilian urban spaces. Its objective is to integrate institutional and individual participants into a shared structure, promoting cooperation in research and collective knowledge construction in Architecture and Urbanism in Brazil and Portuguese-speaking communities. The project focuses on analyzing open space systems (parks and squares) to map landscape architecture in Brazil. Its distinguishing feature is its collaborative ethos, which sets it apart from conventional institutional image databases. By engaging a heterogeneous network of contributors, ARQUIGRAFIA ensures the sustained growth of its digital collection as a dynamic knowledge-building platform. Furthermore, ARQUIGRAFIA seeks to ensure data use and reuse through structure and content standards. This work focuses on this last point.

2.0 Metadata standards and data conversion in legacy systems

ARQUIGRAFIA has developed without the formal adoption of an international metadata standard. This was not limited to the technical domain; rather, it constituted a structural characteristic that affected preservation, interoperability, and data reuse. This condition made it necessary to redesign the platform, situating the discussion within broader debates on legacy data and legacy systems.

Legacy systems may be understood as computational systems developed in earlier technological contexts that remain in operation within organizations. These systems tend to present challenges related to maintenance, integration with contemporary technologies, and limitations in data structuring (Aganette and Santos 2024). Stahlman and Kouper (2024) define legacy data as data systematically collected in the past whose contemporary use and reuse are at risk due to issues such as the absence of adequate metadata, obsolete formats, outdated media, and unsupported software. Their management involves strategies of updating, integration, and system restructuring, accompanied by documentation efforts, evidencing that the issue of legacy systems extends beyond technical obsolescence and involves structural and organizational decisions.

Bisbal et al. (1997) characterize legacy systems as central organizational structures that tend to exhibit rigidity, isolation, and limited capacity for evolution, making modification and adaptation particularly challenging. From this perspective, legacy system evolution and data conversion involve not only technical replacement but also the understanding of

existing structures, semantic mapping, data cleaning, validation of target models, and the maintenance of system continuity throughout the transition process. These challenges include the mapping of existing fields to standardized metadata elements, the normalization of collaboratively produced records, the reinterpretation of previously conflated entities, and the preservation of informational integrity during the coexistence of legacy and redesigned structures.

From a Knowledge Organization perspective, metadata standards may be understood as knowledge organization systems (KOS) that formalize concepts, entities, attributes, and relationships within a given domain (Hjørland 2008). In this sense, the adoption of a metadata standard in ARQUIGRAFIA is not limited to data conversion but also involves a process of conceptual reorganization, making explicit semantic structures that were previously dispersed or implicitly represented in the legacy database.

Martins et al. (2017) identified recurring characteristics associated with legacy systems, organized into seven categories: software structure, maintenance, human resources, obsolete technology, time, financial aspects, and organizational aspects. ARQUIGRAFIA can be situated within the debate on legacy systems, particularly regarding maintenance demands, the need to make its conceptual modeling explicit, and the structural formalization and standardization of its metadata. This situation implies structural challenges, including implicit or dispersed documentation, and difficulties in integration with other systems.

In the context of ARQUIGRAFIA, the adoption of a metadata standard may be interpreted as part of a conceptual and structural data conversion process that requires an incremental strategy. The platform cannot simply reconstruct its database from scratch without risking data loss or discontinuity. Thus, the challenges involved extend beyond the technical dimension and assume a methodological character.

Stahlman and Kouper (2024) emphasize that initiatives aimed at preserving and reusing legacy data require systematic methodologies, analysis of existing practices, and identification of structural problems associated with the data. Among the recurring challenges identified are metadata inconsistency, absence of formal documentation, difficulties in format mapping, and the need for data cleaning and normalization. These elements are directly related to the challenges faced by ARQUIGRAFIA. The adoption of a metadata standard, by explicitly establishing entities and structuring descriptive elements, represents an attempt to address this condition through structured metadata.

Martins, Chervenski, and Bordin (2017) highlight that legacy systems frequently present maintenance problems and lack documentation. In the case of ARQUIGRAFIA, formalizing its model through VRA Core may be understood as a strategy of explicit structural documentation, reducing dependence on tacit knowledge and supporting long-term sustainability.

Stahlman and Kouper (2024) suggest that the perceived value of legacy data and the efforts required for their curation influence institutional decisions regarding data conversion. In this way, the adoption of an international standard such as VRA Core may be viewed as a

strategy for enhancing the value of the collection, expanding its possibilities for reuse and integration with other platforms.

The choice of VRA Core as the standard for ARQUIGRAFIA should not be understood merely as a technical decision but as a strategic intervention in response to the typical characteristics of legacy systems, and it represents an attempt to address issues of interoperability, documentation, and preservation while simultaneously introducing new methodological requirements related to data conversion. Therefore, the redesign of the ARQUIGRAFIA platform is situated within the broader field of legacy system evolution and data conversion, and resonates with contemporary concerns regarding preservation and reuse of legacy data, constituting simultaneously a technical, conceptual, and organizational process that requires in-depth understanding of inherited limitations and strategic planning to overcome them.

3.0 VRA Core as a key: uncovering relationships and reorganizing legacy data

VRA Core has established itself as a metadata standard thanks to the efforts of the professional community managed by the VRA Core Oversight Committee, involving the Visual Resources Association and the Network Development and MARC Standards Office of the Library of Congress (LC). Its original proposal, based on Dublin Core, was adapted to include the description and management of visual resource data. Developed in 1996, and currently in its 4.0 version, released in 2007, VRA Core supports interoperability and the exchange of metadata records (Visual Resources Association 2014).

The standard is structured around three entities – *collection*, *work*, and *image* – whose association is established through the *Relation* element. It is through this architecture that an image can be related to one or more works, a work can be associated with one or more images, and a collection can aggregate records of multiple works or images. With 19 elements – basic metadata units –, VRA Core can be broken down into hierarchically organized subelements: record type (collection/work/image), agent, cultural context, date, description, inscription, location, material, measurements, relationship, rights, source, edition status, style period, subject, technique, text reference, title, and type of work. It also has nine global attributes, applicable to both elements and subelements for the purpose of qualifying them (Visual Resources Association 2014).

From the perspective of Knowledge Organization (KO), VRA Core operates as a Knowledge Organization System (KOS) that formalizes concepts, entities, attributes, and relationships within the cultural heritage domain. The distinction between collection, work, and image is not merely technical, but reflects a conceptual choice about the nature of cultural objects and their representations. This understanding aligns with Dahlberg's (2006) theory, according to which concepts are basic units that express the meaning of things and constitute the core of knowledge organization. In ARQUIGRAFIA, this perspective implies a deliberate effort to define what is meant by work, image, and collection in the context of Brazilian architecture and urban spaces, as well as the relationships among these entities – a process that Hjørland (2008) calls the domain-analytic approach, in which knowledge

structures are understood from the communities and discursive practices in which they are embedded.

Riley's (2009-2010) categorization offers a map that organizes cultural heritage metadata standards according to their strength across four categories – domain, community, function, and purpose. In this map, VRA Core stands out as strongly adhered to *cultural objects* (works of art and architecture) and to *visual resources* of a fixed nature, encompassing *libraries* and *archives* as application communities. Based on this categorization, it is observed that, although the articulation between domain and community already indicates the potential for the incremental development of ARQUIGRAFIA, it is primarily in the function and purpose categories that the standard's relevance to the platform is consolidated.

However, Riley's categorization demands a critical reading in light of the Knowledge Organization literature. As Hjørland (2008) and Dahlberg (2006) argue, KOS cannot be evaluated only by formal and functional characteristics, but also by their epistemological assumptions and domain of application. In this work, it is argued that, in this key, the characterization of VRA Core as weak as a *conceptual model* and *content standard* – as presented by Riley – should not be seen as a limitation, but as a specificity that reflects its original purpose: the description of visual resources and their relationships to the works they represent.

Understanding the function of a metadata standard, as proposed by Riley (2009-2010), implies recognizing its role in the creation and storage of metadata, whether through the definition of specific fields or through guidance on the filling of values. In this scope, Riley classifies VRA Core as a *record format* and *structure standard*, being considered semi-weak as a *controlled vocabulary* and weak as a *conceptual model* and *content standard*. According to Hodge (2000), KOS can be grouped into three categories – term lists, categorization and classification systems, and conceptual models and ontologies – and VRA Core transits among them: as a *record format*, it operates in information structuring; as a *structure standard*, it defines the elements necessary to describe a resource; as a *controlled vocabulary*, it guides term selection; and, although less formalized, it offers a *conceptual model* by defining entities and relationships.

This hybridity, often interpreted by Riley (2009-2010) as a weakness, can be read, from a KO perspective, as flexibility and adaptability to different descriptive contexts. The strength of VRA Core as a *record format* stems from the specific encodings adopted in its set of elements, which underpins the discussion on the format to be used in ARQUIGRAFIA. As a *structure standard*, it defines the elements applicable to certain materials or purposes, although there is no direct correlation between structures and formats – an aspect that ARQUIGRAFIA addresses through a conceptual study for the definition of data types and dictionaries.

The classification of VRA Core as semi-weak as a *controlled vocabulary* and weak as a *conceptual model* and *content standard*, in Riley's (2009-2010) terms, demands a deeper reflection on the role of these dimensions. Hjørland (2008) argues that the effectiveness of a KOS is not measured only by formal criteria, but by its adequacy to the domain and to

users' needs. In this study, it is considered that, in ARQUIGRAFIA, the weakness as a *controlled vocabulary* can be compensated by the adoption of established thesauri, while its weakness as a *conceptual model* signals the need to make explicit semantic relationships between entities – a task that can be guided by principles of semantic mapping and interoperability among KOS, as discussed by Yuan et al. (2025) based on the ISO 25964 and ISO 11179 standards. This characterization was not taken as a demerit, but as a reinforcement of the usefulness of *controlled vocabularies* in specific domains and of the contribution of the *conceptual model* to the understanding of relationships among ARQUIGRAFIA's informational resources.

As for purpose, Riley (2009-2010) highlights that VRA Core acts primarily as a *descriptive standard*. In this study, it is argued that, in ARQUIGRAFIA, this function translates into the recording of data that enables searching, discovering, and interpreting informational resources on architectural works and their representations. As an adjacent purpose, VRA Core also acts as a *rights standard* (by ensuring access or establishing use restrictions) and as a *technical standard* (by recording actions performed on resources over time).

Finally, the conceptual dimension of VRA Core in ARQUIGRAFIA gains density by engaging with the theoretical approaches distinguished by Hjørland (2008): the concept-based approach, the domain-analytic approach, and the classification-based approach. In this study, it is observed that the choice of the standard reflects this plurality: (1) by adopting an entity- and attribute-based model, it aligns with the concept-based approach; (2) by considering the domain of architecture and urban spaces, it approximates the domain-analytic approach; and (3) by structuring data according to a predefined scheme, it operates within a classificatory logic. This multiplicity, far from being a problem, reveals the complexity of knowledge organization in collaborative digital environments and points to the need for an integrative approach – such as that proposed by Dahlberg (2006) – that encompasses the conceptual, semantic, and pragmatic dimensions.

In summary, the adoption of VRA Core in ARQUIGRAFIA transcends its technical dimension and is inscribed in a broader project of knowledge organization, involving conceptual, epistemological, and pragmatic choices. By establishing entities, attributes, and relationships, the standard operates as a KOS that organizes not only the data but also the practices of description, retrieval, and interpretation of information. The KO literature – represented here by authors such as Dahlberg (2006), Hjørland (2008), Hodge (2000), and Riley (2009-2010) – offers a consistent framework for critically evaluating these choices, evidencing that the effectiveness of a metadata standard lies not only in its formal characteristics but in its adequacy to the domain, to communities of practice, and to the goals of knowledge communication and preservation. The ARQUIGRAFIA experience with VRA Core, by articulating technical and conceptual dimensions, contributes to the contemporary debate on knowledge organization in collaborative digital environments, reinforcing the importance of a theoretically informed approach to the design and implementation of information systems in the cultural heritage domain.

4.0 From standard to platform: the application of VRA Core in ARQUIGRAFIA

The VRA Core model was used both to define the structure of database tables and to determine the values assigned to their fields. As a result, the system's database tables were divided into two groups: those required to represent the VRA Core model and those needed for the system's internal operation. To distinguish them, tables corresponding to VRA Core elements were assigned the prefix "vrac_". For example, the Title element corresponds to a table named vrac_titles. In this way, tables were created and associated with each of the 19 elements of the model. The following technologies were adopted: PHP (version 8.4) as the programming language, Laravel as the framework (version 12), and MySQL (version 8) for data management.

The legacy database was analyzed to account for the content that had been entered by users, who are not trained information systems professionals. So, the labels and contents of the database tables frequently did not correspond to one another or follow a consistent standard. This analysis guided discussions within the development team in establishing the conceptual foundations of the project aimed at achieving a comprehensive understanding of the relationship between the VRA Core standard and the legacy information system. A new database model was designed, and following a series of revisions, it was implemented with adaptations developed to address ARQUIGRAFIA's specific requirements and context.

The main adaptations made to the VRA Core model are described as follows. As a general measure applied to all elements, the *snake_case* naming convention was adopted, using lowercase plural forms for table names and lowercase singular forms for field names. Additionally, an **id** field was added as the primary key, following the **UUID** standard, along with the **created_at** and **updated_at** fields, which are automatically created and managed by Laravel. Regarding specific elements, the Relation element was removed because relationships between entities in Laravel are implemented through dedicated relational tables, making the original table unnecessary. For the Agent element, some fields were separated into additional tables – namely **vrac_agent_dates**, **vrac_agent_roles**, and **vrac_contributor_names** – to improve vocabulary control and database organization. Finally, the Location element was not adopted, as it would not adequately support the storage and retrieval of georeferenced data; instead, a system-specific table was created.

To provide controlled values for some VRA Core element fields, the Vocabulário Colaborativo de Artes e Arquitetura (VCAA) thesaurus was adopted. Most VRA Core elements were mapped to specific controlled lists from VCAA: vrac_agent_roles was linked to "Pessoas, Personagens e Profissões"; vrac_materials to "Materiais"; vrac_style_periods to "Estilos (arquitetura)"; vrac_techniques to "Técnicas"; vrac_worktypes to "Função"; and vrac_subjects to the entire vocabulary. However, two elements – vrac_contributor_names and vrac_cultural_contexts – were not derived from the VCAA vocabulary.

For the data migration process, forty-one legacy tables were analyzed. Metadata describing the ARQUIGRAFIA collection were identified in seven tables: authors, comments, photo_author, photos, suggestions, institutions, and tags. However, the comments and suggestions tables were discarded, as their analysis did not yield satisfactory results for data conversion. Analysis of the data in the authors, photos, and tags tables revealed significant inconsistencies. In the authors and photos tables, problems involved the name,

imageAuthor, and project_author fields: the name field inconsistently recorded both individuals and institutions; project_author was largely empty; and imageAuthor referred to custodians rather than producers. In the tags table, the name field contained numbers, dates, and free-text expressions, often with multiple terms and different spellings within a single field. These inconsistencies are likely related to the collaborative nature of the platform, which allows contributions from users with varying indexing expertise. Based on the observations described above, it was possible to establish correspondences between the legacy data and the VRA Core model, as shown in Table 1.

Table 1 Matches between legacy data and VRA Core metadata elements

Legacy Table	Legacy Fields	VRA Core	Processing
Tags	name	Subject	Normalized to avoid duplication of terms
Fotos	allowCommercialUses allowModifications	Rights	Unified to create a new standard model
Fotos	city country district state street	Location	Unified to enable georeferencing
Fotos	dataCriacao hygieneDate workdate	Date	Converted from text to ISO 8601 format
Fotos	name	Title	
Fotos	description	Description	
Fotos	imageAuthor	Agent	
Fotos	aditionalImageComments	State Edition	
Fotos	collection	Location	
Authors	name	Agent	
institutions	name	Agent	

The launch of the platform will make it possible to empirically evaluate the impacts of the proposed restructuring, but several tangible benefits are expected from the adoption of the VRA Core standard and the corresponding reengineering of the legacy system. First, the new data model promotes greater standardization of metadata fields and input options, contributing to improved data consistency and quality. Second, the introduction of more specific and semantically defined fields is expected to facilitate the content submission process for non-specialist users, reducing ambiguities and classification errors. This addresses issues identified in the legacy database, such as the former "Author" field, which combined distinct entities including architects, photographers, and institutions. Finally,

the alignment with an internationally recognized metadata standard enhances interoperability, enabling future integration with external repositories and information systems, both for data exchange and for the incorporation of new information into ARQUIGRAFIA itself. These improvements are expected to increase the usability, scalability, and long-term sustainability of the platform. Taken together, the mapping process revealed that the migration of legacy data to VRA Core involved not only technical correspondences between fields but also the reinterpretation and normalization of heterogeneous records. The identified inconsistencies reinforced the need for a structured metadata model capable of making entities and relationships more explicit, thereby supporting greater consistency, interoperability, and long-term sustainability of the platform.

5.0 Conclusions

When describing an image of a building, a monument, or a specific urban space, it is necessary to provide complete and accurate information about the original work or place. It is important not to confuse the metadata of the original work with the metadata of its image, since the image stands in for the original for those who cannot see it directly.

The use of VRA Core as a metadata standard also works as a mediation tool between ARQUIGRAFIA's collection and its users, since ARQUIGRAFIA is a collaborative environment where anyone can upload images. By introducing a clear distinction between entities such as work, image, and collection, VRA Core provides a conceptual framework capable of reorganizing heterogeneous data and clarifying relationships between resources, whether legacy data or new records added by platform users.

This restructuring contributes to improving the consistency of descriptive data and establishes a more stable foundation for the long-term management of the platform's information resources.

At the same time, the transition from a legacy data structure to a model aligned with VRA Core revealed relevant methodological and curatorial challenges. Processes such as semantic mapping, normalization of descriptive values, and separation of previously conflated entities required a critical reinterpretation of legacy records and careful management of data conversion. The experience shows that adopting a metadata standard is not merely a technical operation but a process that involves conceptual revision, data curation, and systematic documentation to ensure the integrity and sustainability of digital repositories. Despite these advances, the adoption of VRA Core does not eliminate the need for ongoing curation and validation, since inconsistencies inherited from legacy data and the collaborative nature of the platform continue to require human interpretation and management.

Taken together, these observations show that the adoption of VRA Core in ARQUIGRAFIA not only guided the structural redesign of the database but also functioned, from a Knowledge Organization perspective, as a conceptual framework that revealed implicit structures in legacy data and enabled the reinterpretation of relationships among works, images, and collections, supporting the systematic reorganization of knowledge on the platform.

Generative artificial intelligence tools, specifically ChatGPT and DeepSeek, were used exclusively to support the editorial revision of this manuscript and to assist in the preparation of its English-language version.

References

Aganette, Elisângela Cristina, and Rodrigo Fernandes Santos. 2024. "Inovação em BPMS: O impacto das plataformas low-code nas organizações." In Anais do XXIV Encontro Nacional de Pesquisa em Ciência da Informação (ENANCIB), GT-8 – Informação e Tecnologia. Vitória, ES: ANCIB/PPGCI-UFES. https://enancib.emnuvens.com.br/enancib/en/article/view/873.

Bisbal, Jesus, Deirdre Lawless, Bing Wu, Jane Grimson, Vincent Wade, Ray Richardson, and Donie O'Sullivan. 1997. "An Overview of Legacy Information System Migration." In Proceedings of the Joint 1997 Asia Pacific Software Engineering Conference and International Computer Science Conference (APSEC'97/ICSC'97), 529–530. Hong Kong: IEEE Computer Society. https://doi.org/10.1109/APSEC.1997.640219.

Dahlberg, Ingetraut. 2006. "Knowledge Organization: A New Science?" Knowledge Organization 33, no. 1: 11–19. https://doi.org/10.5771/0943-7444-2006-1-11.

Hjørland, Birger. 2008. "What Is Knowledge Organization (KO)?" Knowledge Organization 35, no. 2–3: 86–101. https://doi.org/10.5771/0943-7444-2008-2-3-86.

Hodge, Gail. 2000. Systems of Knowledge Organization for Digital Libraries: Beyond Traditional Authority Files. Washington, DC: Digital Library Federation/Council on Library and Information Resources. https://www.clir.org/wp-content/uploads/sites/6/pub91.pdf.

Martins, Daniele, Alex Chervenski, and Andréa S. Bordin. 2017. "Identificação de características de sistemas legados a partir da análise de conteúdo da literatura." In Anais da I Escola Regional de Engenharia de Software (ERES 2017), 81–88. Porto Alegre: Sociedade Brasileira de Computação. https://sol.sbc.org.br/index.php/eres/article/view/10084.

Riley, Jenn. 2009–2010. Seeing Standards: A Visualization of the Metadata Universe. Bloomington, IN: Indiana University Libraries. https://jennriley.com/wp-content/uploads/2025/07/seeingstandards.pdf.

Stahlman, Gretchen, and Inna Kouper. 2024. "Navigating the Challenges of Legacy Data: A Systematic Literature Analysis." Proceedings of the ALISE Annual Conference. https://doi.org/10.21900/j.alise.2024.1751.

Visual Resources Association. 2014. “An Introduction to VRA Core.” Library of Congress. https://www.loc.gov/standards/vracore/VRA_Core4_Intro.pdf.

Yuan, Jingshu, Kexin Zhai, Hongxin Li, and Man Yuan. 2025. “Research on the Construction and Mapping Model of Knowledge Organization System Driven by Standards.” Computer Standards & Interfaces 92: 103905. https://doi.org/10.1016/j.csi.2024.103905