RiiTE
Nº 20 Junio 2026 pp. 67-88 ISSN: 2529-9638 DOI: https://doi.org/10.6018/riite.679231

Systematic review of criteria for digital resource catalogues in language learning: foundations for a Catalan catalogue

Digital resources for learning Catalan. A systematic review of the characteristics and requirements

Mònica Vilosa Sánchez

Departament d'Educació i Formació Professional, Generalitat de Catalunya (España)
mvilosa@xtec.cat

Juan González-Martínez

Universitat de Girona (España)
juan.gonzalez@udg.edu

Recibido: 10/09/2025
Aceptado: 11/01/2026
Publicado: 01/06/2026

ABSTRACT

Introduction: The proliferation of digital resources for language learning has increased the need for catalogues that facilitate the search, selection and pedagogical use of materials, especially in languages with non-hegemonic digital ecosystems such as Catalan. Method: A systematic review was conducted following PRISMA 2020. Scopus and Web of Science were consulted using two search strings focused on digital resources, repositories, classification and language learning. After removing duplicates, inclusion and exclusion criteria were applied to peer-reviewed articles in English or Spanish with full-text access. The screening was documented with a PRISMA diagram, and 73 studies were included. The synthesis was based on qualitative coding and thematic analysis. Results: The findings were organised into five criteria axes: (1) classification and categorisation, with a predominance of filters by language skills and CEFR levels; (2) metadata, highlighting the need for pedagogical and linguistic fields and interoperability through standards; (3) pedagogical quality, with proposals for expert validation and automated verification of descriptors; (4) architecture and openness, with an emphasis on open licences and protocols such as OAI-PMH; and (5) support for mixed contexts, incorporating information on autonomy, usage guidelines and the integration of collaborative tools. Conclusions: The review proposes an integrated framework of transferable criteria for designing and evaluating language resource catalogues and provides the basis for the development of a catalogue in Catalan. It is recommended that the criteria be empirically validated with teachers and learners and that future reviews be extended to additional sources and languages.

KEYWORDS

Language learning; Digital resource catalogues; Educational repositories; Metadata; Catalan

RESUMEN

Introducción: La proliferación de recursos digitales para el aprendizaje de lenguas ha aumentado la necesidad de catálogos que faciliten la búsqueda, la selección y el uso pedagógico de materiales, especialmente en lenguas con ecosistemas digitales no hegemónicos como el catalán. Método: Se realizó una revisión sistemática siguiendo PRISMA 2020. Se consultaron Scopus y Web of Science mediante dos cadenas de búsqueda centradas en recursos digitales, repositorios, clasificación y aprendizaje de lenguas. Tras eliminar duplicados, se aplicaron criterios de

inclusión y exclusión a artículos revisados por pares, en inglés o español y con acceso a texto completo. El cribado se documentó con un diagrama PRISMA y se incluyeron 73 estudios. La síntesis se basó en codificación cualitativa y análisis temático. Resultados: Los hallazgos se organizaron en cinco ejes de criterios: (1) clasificación y categorización, con predominio de filtros por habilidades lingüísticas y niveles del MCER; (2) metadatos, destacando la necesidad de campos pedagógicos y lingüísticos y la interoperabilidad mediante estándares; (3) calidad pedagógica, con propuestas de validación experta y verificación automatizada de descriptores; (4) arquitectura y apertura, con énfasis en licencias abiertas y protocolos como OAI-PMH; y (5) soporte para contextos mixtos, incorporando información sobre autonomía, orientaciones de uso e integración de herramientas colaborativas. Conclusiones: La revisión propone un marco integrado de criterios transferible para diseñar y evaluar catálogos de recursos lingüísticos y fundamenta el desarrollo de un catálogo en catalán. Se recomienda validar empíricamente los criterios con docentes y aprendientes y ampliar futuras revisiones a fuentes y lenguas adicionales.

PALABRAS CLAVE

Aprendizaje de lenguas; Catálogos de recursos digitales; Repositorios educativos; Metadatos; Catalán

CITA RECOMENDADA:

Vilosa, M. y González-Martínez, J. (2026). Digital resources for learning Catalan. A systematic review of the characteristics and requirements. *RiiTE Revista interuniversitaria de investigación en Tecnología Educativa*, *20*, 67-88. https://doi.org/10.6018/riite.679231

Principales aportaciones del artículo y futuras líneas de investigación:

- The literature review isolates the technical characteristics necessary in a repository of educational resources for language learning.

1. INTRODUCTION

The incorporation of digital resources into language teaching and learning has intensified significantly in recent years, both in formal education contexts and in autonomous learning situations. This increase is a response to the progressive digitisation of educational practices and the need to offer flexible responses to the diversity of learners' profiles, rhythms and needs (Churchill, 2017; Kieu et al., 2021). In this context, mobile-assisted language learning (MALL) has established itself as a key strategy for eliminating time and space constraints, allowing on-demand access to materials (Zervas & Sampson, 2010; Zervas & Sampson, 2013).

In this scenario, digital resources can contribute significantly to improving student motivation, involvement and commitment, provided they are integrated into coherent teaching designs geared towards clear learning objectives. However, the growing availability of digital materials does not in itself guarantee an improvement in educational quality. Teachers and learners are often faced with a large volume of scattered, heterogeneous resources of varying quality. This information overload makes it essential to develop semi-automatic metadata generation systems to manage the explosion of data (Park & Brenza, 2015).

In view of this situation, digital resource repositories and catalogues are emerging as key tools for facilitating the organisation, search and selection of educational materials. However, the literature

shows that many of these digital spaces have shortcomings in terms of pedagogical systematisation, metadata quality and the definition of clear criteria to guide the educational use of resources (Muñoz & Román-Mendoza, 2018; Zervas & Sampson, 2013). The lack of standardisation often means that the potential of linguistic archives remains latent or underutilised (Burke & Zavalina, 2020).

In the field of language learning, these limitations become particularly relevant, given the need to take into account specific dimensions such as linguistic skills, levels of competence, communicative objectives and contexts of use. Furthermore, the coexistence of formal education and autonomous learning contexts reinforces the importance of having catalogues that allow for flexible adaptation of resources.

Based on this approach, the present article aims to identify and systematise criteria for digital resource catalogues intended for language learning, based on a systematic review of the scientific literature. This contribution aims to provide a sound basis for guiding both future research and initiatives for the design and improvement of resource catalogues, with a special focus on languages with non-hegemonic digital ecosystems, such as Catalan (European Language Resources Association, 2024; Department of Education of the Government of Catalonia, 2022).

2. THEORETICAL FRAMEWORK

2.1. Digital resources and language learning in current educational contexts

In recent years, the use of digital resources in language teaching and learning has experienced remarkable growth, both in formal education contexts and in autonomous learning environments. This expansion responds, on the one hand, to the progressive digitisation of educational practices and, on the other, to the need to cater to the diversity of learners' profiles, rhythms and needs (Churchill, 2017; Kieu et al., 2021). The evolution towards "Web 2.0" has allowed users not only to consume content, but also to create it, promoting constructivist approaches (Kárpáti, 2009).

The specialist literature points out that digital resources can contribute significantly to increasing student motivation, involvement and commitment, provided that they are integrated into coherent teaching designs geared towards clear learning objectives. Elements such as gamification and digital games have been validated as tools that reduce anxiety and encourage the "willingness to communicate" (Reinders, 2017). These resources can facilitate content presentation, guided practice, immediate feedback and self-assessment, aspects that are particularly relevant in language learning.

From a pedagogical perspective, digital resources should not be understood as isolated elements, but as tools integrated into an overall teaching design. Their potential lies not only in their technological characteristics, but also in the way they are linked to learning objectives, proposed activities and the educational context. In this sense, the literature highlights the importance of a learner-centred approach, which allows resources to be adapted to their interests, language proficiency level and specific needs (Bin-Hady & Al-Tamimi, 2021).

This approach is particularly relevant in blended learning contexts, where formal, teacher-guided teaching situations coexist with autonomous or semi-autonomous learning processes. In these environments, digital resources can act as a bridge between the classroom and learning outside the

educational centre, expanding opportunities for exposure to the language and its use in diverse communicative contexts.

However, the quantitative increase in available digital resources does not in itself guarantee an improvement in educational quality. Several studies warn of the risk of dispersion and information overload, especially for teachers, who often have to spend a considerable amount of time searching for, selecting and evaluating resources that are suitable for their teaching practice. In this scenario, the key issue is not so much the availability of resources as the existence of mechanisms that facilitate their organisation, search and pedagogical use.

2.2. Repositories and catalogues of digital resources for language learning

Given the proliferation of digital resources, repositories and catalogues have become established as fundamental tools for the management, organisation and dissemination of digital educational materials. Although they are often used interchangeably, these two concepts serve different functions (Muñoz & Román-Mendoza, 2018; Anyaoku et al., 2019). While repositories act as access warehouses, catalogues add a layer of pedagogical selection and filtering.

Digital educational resource repositories are mainly defined as spaces for storing and accessing digital materials, often under open licences, which allow for the reuse and sharing of content. In the field of language learning, initiatives linked to open educational resources (OER) and learning object repositories (LOR) stand out, offering materials for different languages, levels and educational contexts. Federations such as CLARIN and META-SHARE exemplify the trend towards the creation of federated infrastructures that enable the exchange of documented linguistic resources (Hinrichs & Krauwer, 2014; Piperidis, 2012).

Resource catalogues, on the other hand, go beyond the function of a container. In addition to hosting resources, they incorporate classification systems, descriptors and criteria that allow users to search, filter and select materials according to specific pedagogical needs. The added value of the catalogue thus lies in its ability to transform a collection of resources into an effective tool for supporting teaching decisions and autonomous learning. Content curation is essential here to prevent materials from becoming obsolete (Hernández Muñoz & Román-Mendoza, 2018).

The literature shows that, in the field of language teaching, the most useful catalogues are those that integrate criteria related to language skills, proficiency levels, learning objectives and types of activities. These elements facilitate the identification of appropriate resources for each teaching situation and contribute to the personalisation of learning (Kárpáti, 2009).

In addition, repositories and catalogues can encourage the creation of communities of practice, promoting collaboration between teachers and the sharing of materials, experiences and good practices. In this sense, some spaces incorporate peer review or evaluation systems that contribute to improving the quality of available resources.

Despite these advances, the literature also points to significant limitations. Not all repositories have clear classification criteria or sufficiently detailed metadata, which makes it difficult to retrieve resources efficiently. In addition, many catalogues are not specifically designed for language learning and have a generic orientation that does not take into account the particularities of this field.

2.3. Classification and categorisation systems for digital resources

The classification of digital resources is a central element in any catalogue aimed at facilitating their pedagogical use. In the field of language learning, the literature shows that the availability of resources is not sufficient if they are not organised according to criteria that respond to the real needs of teachers and learners (Zervas & Sampson, 2013; Piperidis, 2012). The use of knowledge maps and ontologies enables semantic searches that go beyond the simple pairing of keywords (Yan & Xiao-yue, 2007; Winslow et al., 2019).

Classification systems allow resources to be structured according to relevant dimensions, such as language proficiency level, communication skills involved, learning objectives, or type of activity proposed. This organisation facilitates the efficient retrieval of materials and helps to reduce the time teachers have to spend searching for and selecting resources.

The literature highlights that the most useful classifications are those that integrate pedagogical and linguistic criteria in a combined manner. This includes classification according to linguistic skills, as well as according to grammatical, lexical, or pragmatic aspects. The use of standard application profiles such as IEEE LOM allows general categories to be adapted to the specific needs of the MECR (Tobar & De Lima, 2007). These criteria allow resources to be selected that are aligned with the curriculum objectives and the needs of the students.

In addition, several studies advocate the need for flexible classification systems that allow for the combination of multiple criteria and avoid rigid structures. In this sense, dynamic categorisation models based on multiple tags or descriptors (folksonomies) facilitate more accurate searches and better adaptation to diverse educational contexts, especially when they are based on controlled vocabularies to ensure consistency (Matthews et al., 2010).

In blended learning environments, classification takes on special relevance, as it must respond to both teaching use in formal contexts and autonomous learning. Criteria such as the degree of autonomy required, the type of feedback, or the possibility of use outside the classroom become particularly significant in this context.

2.4. Metadata, standards and retrieval of educational resources

Metadata is a key element in the functioning of digital resource repositories and catalogues, as it allows materials to be described, organised and retrieved efficiently. In the field of language learning, the quality and structure of metadata directly determine the possibility of pedagogical reuse of resources (Zervas & Sampson, 2010; Hinrichs & Krauwer, 2014).

From a functional perspective, metadata provides essential information about resources, such as content, format, educational level, skills involved, and context of use. This information allows users to quickly identify the relevance of resources for a specific educational purpose. The use of unique identifiers such as ISLRN ensures the visibility and citability of these resources (Mapelli et al., 2018).

The literature highlights the importance of using recognised metadata standards, such as Dublin Core or IEEE LOM, which guarantee interoperability between different repositories and systems (Bourda & Doan, 2003). However, several authors point out that the application of general standards may be insufficient if it is not adapted to the particularities of language learning.

In this sense, there is a clear need to incorporate specific metadata related to language skills, proficiency levels, communication objectives and methodological approaches. This type of information allows for more accurate retrieval of resources and contributes to the personalisation of learning. To reduce the costs of manual creation, automatic extraction tools based on natural language processing are being explored (Casali et al., 2014).

With regard to retrieval mechanisms, the literature indicates that search systems based exclusively on keywords are often insufficient. For this reason, approaches have been developed that combine structured metadata, controlled vocabularies, and other techniques aimed at improving the accuracy and relevance of results. Among these approaches, solutions based on the semantic web have been proposed to improve the retrieval of educational resources (Bourda & Doan, 2003).

2.5. Quality, pedagogical criteria and use of catalogues

The quality of digital resources and the catalogues that organise them is a key factor in ensuring their pedagogical usefulness. In language teaching, the literature emphasises that the presence of explicit quality criteria facilitates the selection of appropriate materials and contributes to improving the effectiveness of teaching practices (Krajcso, 2016; Piedra et al., 2015; González-Blanco et al., 2017; Del et al., 2016).

Quality criteria should not be understood as closed normative lists, but rather as guidelines that help to assess the relevance, reliability and pedagogical suitability of resources. Good practice in the digital age is no longer defined solely by its replicability, but also by its functionality, connectivity and capacity for social transformation (Hernández Muñoz & Román-Mendoza, 2018). Among the most notable aspects are consistency with learning objectives, clarity in the presentation of content, suitability for the level of linguistic competence, and the possibility of providing meaningful feedback.

In the case of catalogues, quality depends not only on individual resources, but also on the criteria used to describe, classify and present them to users. Catalogues with explicit pedagogical criteria facilitate comparison between resources and support a more conscious and informed selection.

The literature shows that incorporating quality criteria into catalogues can encourage reflective practices on the part of teachers and promote the creation of communities of practice aimed at the continuous improvement of materials.

2.6. The case of Catalan: specific needs and opportunities in the digital sphere

The learning of Catalan takes place within a specific sociolinguistic and educational context that presents opportunities and challenges in the field of digital resources. Despite the existence of institutional initiatives and an active digital ecosystem, difficulties persist in relation to the systematisation, organisation and pedagogic tion of available resources (European Language Resources Association, 2024; Department of Education of the Government of Catalonia, 2022). Low-density languages or small digital ecosystems require data collection strategies using crawlers to increase their visibility (Baldwin et al., 2006).

This situation is particularly relevant in a scenario characterised by the diversity of learner profiles, which includes both school pupils and adults in autonomous learning processes. In this context, the availability

of resources is not sufficient if it is not accompanied by criteria that facilitate access, selection and adaptation to the specific needs of users. Initiatives such as the TL Plan in Spain seek precisely to transform public documentation into digital linguistic resources to promote the language (Moreno et al., 2019).

From a language planning perspective, Catalan shares challenges with other languages that do not have as large a digital education market as the hegemonic languages. The literature shows that, in these cases, the definition of clear criteria for the classification, description and quality of resources takes on particular relevance in ensuring the digital survival of the language (Mohamad et al., 2011).

2.7. Summary of the theoretical framework and connection with the research objectives

The theoretical framework developed has made it possible to identify the main elements that determine the pedagogical usefulness of digital resource catalogues for language learning. The literature reviewed highlights the importance of having clear classification criteria, adequate metadata and pedagogical quality criteria, especially in blended learning contexts (Muñoz & Román-Mendoza, 2018; Zervas & Sampson, 2013).

This synthesis reveals that an effective model must manage the entire life cycle of the language resource: from acquisition (through crawling or donations from speaker communities), through manufacturing (validation and annotation by experts or crowdsourcing), to dissemination (publication through linked data structures) (Gonzalez et al., 2017).

Therefore, the gap identified is not only one of availability, but also of information architecture: systems are needed that separate structural representation (metadata) from content representation to enable true personalisation of learning (Agosti et al., 2011; Slimani et al., 2016).

3. METHODOLOGY

3.1. Research design

This study is part of a qualitative research paradigm and adopts systematic literature review as its methodological strategy, with the aim of identifying and systematising criteria for catalogues of digital resources for language learning. The review was carried out following the guidelines of the PRISMA 2020 statement, ensuring the transparency and replicability of the process (Page et al., 2021). The methodological guidelines established by Brereton et al. (2007) and Kitchenham (2004) were followed for the formulation of research questions and the elimination of biases.

3.2. Information sources and search strategy

The bibliographic search was carried out in two international reference databases: Scopus and Web of Science. These databases were selected due to their broad multidisciplinary scope and the fact that they index high-impact scientific journals in the fields of education, educational technology and language sciences.

The search process was carried out iteratively. In the initial phase, different combinations of keywords were explored with the aim of identifying terms that produced relevant results aligned with the research

objectives. This process allowed the search strategy to be progressively refined and the final search strings to be defined.

Finally, two main search strings were established: - language AND digital AND resources AND repository - language AND digital AND resources AND classification

In the case of Scopus, the second chain generated a large number of results. For this reason, the open access filter was applied in order to guarantee the availability of the full text of the documents and facilitate their subsequent analysis. In the case of Web of Science, searches were carried out on the title, abstract and keyword fields.

The searches were limited to documents published in English or Spanish, without applying specific time restrictions, in order to obtain a broad overview of the evolution of studies on catalogues and repositories of digital resources for language learning.

3.3. Inclusion and exclusion criteria

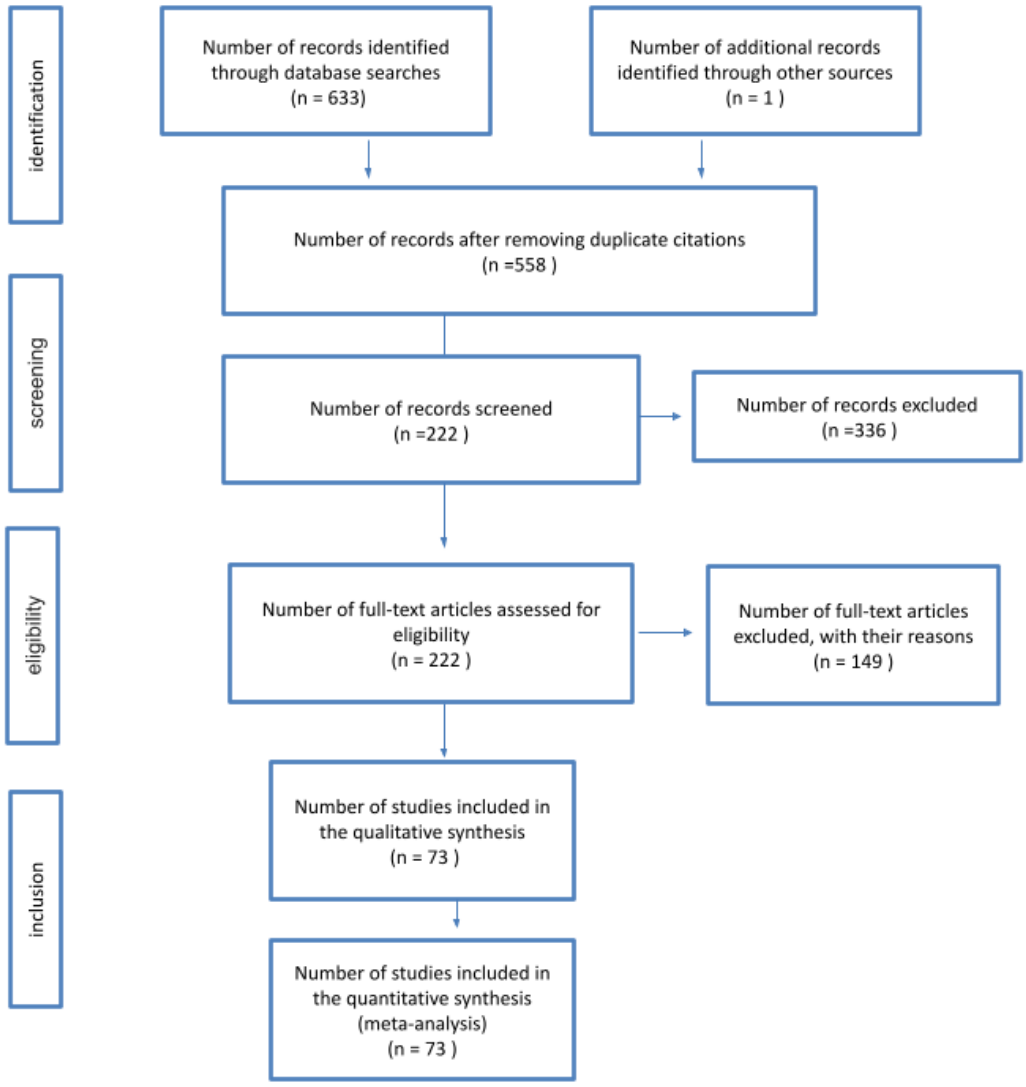
In accordance with the PRISMA model, studies that met the following criteria were included: peer-reviewed scientific articles; studies focused on repositories, catalogues or systems for organising digital resources for language learning; publications in English or Spanish; documents with access to the full text.

Studies were excluded that: addressed areas not related to education or language learning; focused exclusively on digital libraries or institutional repositories without an explicit pedagogical orientation; did not provide relevant information on classification systems, metadata structures or quality criteria applicable to digital resources.

3.4. Study selection process

In accordance with the PRISMA phases, 633 records were initially identified, reduced to 558 after eliminating duplicates. In the screening phase, 336 records were excluded based on title and abstract, selecting 222 articles for full-text evaluation. After a thorough reading, 148 additional articles were discarded due to lack of pedagogical or thematic relevance. Finally, 73 studies constituted the final sample for qualitative synthesis (Figure 1).

Figure 1
Flow chart of the methodological process followed



3.5. Analysis of the information

The studies included in the final sample were imported into NVivo software for qualitative analysis. The analysis was carried out using a thematic coding process that combined a deductive approach, based on the research objectives and questions, with an inductive approach aimed at identifying emerging categories from the data.

The categories defined were related to the classification of resources, metadata, quality criteria, types of repositories and the pedagogical use of catalogues. This coding process made it possible to identify recurring patterns and establish relationships between the criteria described in the literature.

3.6. Synthesis and presentation of results

Based on the qualitative analysis, the results were synthesised in the form of systematised criteria. This strategy allows us to respond directly to the research objectives and establishes a sound basis for the discussion and conclusions of the study.

4. RESULTS

The results of the systematic review are presented as a systematisation of criteria derived from the qualitative analysis of the 73 studies included in the synthesis. In accordance with the research objectives, this section is organised as a set of structured criteria that inform the design and use of digital resource catalogues for language learning.

The categories presented below emerge from the thematic coding process carried out with NVivo and are aligned with the main axes identified in the literature: classification of resources, metadata and description systems, pedagogical quality, repository architecture, and support for blended learning contexts.

4.1. Criteria related to the classification and categorisation of resources

The classification of digital resources is a central element in ensuring their pedagogical usefulness, integrating linguistic and didactic criteria (Zervas & Sampson, 2013; Di Martino, 2009). A recurring criterion is organisation according to language skills (oral and written comprehension/expression) and subsystems such as vocabulary and grammar (Krajcso, 2016; Bin-Hady & Al-Tamimi, 2021).

It is also essential to link the materials to the competence levels of the Common European Framework of Reference for Languages (CEFR). In the field of MALL (Mobile Assisted Language Learning), categorisation must take into account the technical characteristics of mobile devices (Zervas & Sampson, 2010). The most effective systems allow for the use of multiple criteria and avoid rigid structures (table 1).

Table 1
Most frequent classification criteria in the studies included (n=73)

Classification criteria	Frequency (studies)	Examples of sources
Language skills (4 languages)	22/73 (30%)	Krajcso (2016), Bin-Hady & Al-Tamimi (2021)
CEFR levels (A1-C2)	18/73 (24%)	Zervas & Sampson (2010, 2013), Tobar & De Lima (2007)
MALL/mobile devices	12/73 (16%)	Zervas & Sampson (2010), Park & Brenza (2015)
Flexible folksonomies	9/73 (12%)	Matthews et al. (2010)
Semantic ontologies	6/73 (8%)	Winslow et al. (2019)

Specific example - Cambridge model: Cambridge filters resources by "Qualification/CEFR/Skill/Resource Type/Age Level". This multi-level navigation model can be replicated for Catalan catalogues (Krajcso, 2016).

4.1.1. Language skills

Organisation according to the four basic skills (listening comprehension, speaking, reading comprehension, writing) appears in 22/73 studies (30%). This categorisation is particularly useful for autonomous learning, as it allows users to select resources aimed at developing specific skills (Krajcso, 2016; Bin-Hady & Al-Tamimi, 2021). Furthermore, the breakdown by subsystems (phonetics, lexicon, grammar, pragmatics) adds a second layer of precision (Zervas & Sampson, 2013).

4.1.2. MECR proficiency levels

18/73 studies (24%) incorporate the A1-C2 level classification of the Common European Framework of Reference. This standardisation facilitates alignment with national curricula and comparability between resources (Tobar & De Lima, 2007; Krajcso, 2016). In the case of Catalan, the adoption of the MECR classification would ensure interoperability with other European catalogues (Piedra et al., 2015).

4.1.3. Classification by MALL contexts

12/73 studies (16%) point to the need for specific criteria for resources on mobile devices. The criteria include: screen size, optimal session duration (microlearning), connection type (online/offline) and multi-platform compatibility (Zervas & Sampson, 2010; Park & Brenza, 2015).

4.2. Criteria related to metadata and resource description

Structured metadata is essential for the pedagogical retrieval of resources (Zervas & Sampson, 2010; Hinrichs & Krauwer, 2014; Piperidis, 2012). Basic descriptors (educational level, format) and specific metadata such as communication objectives and methodological approach should be included. The use of standards such as IEEE LOM facilitates interoperability, although adapted application profiles are often needed (Park & Brenza, 2015; Bourda & Doan, 2003).

Quantitative analysis NVivo: 42/73 studies (57%) advocating metadata standards. Distribution by type (table 2).

Table 2
Metadata standards used in repositories (n=73 studies)

Metadata standard	% of studies	Studies (n=73)	Main sources
IEEE LOM	28%	21	(Hinrichs & Krauwer, 2014; Park & Brenza, 2015; Tobar & De Lima, 2007)
Dublin Core	19	14	(Bourda & Doan, 2003; Piperidis, 2012)
OAI-PMH (interoperability)	15	11	(Karadia & Sahoo, 2021; Piedra et al., 2015)
CLARIN-compatible	8	6	(Hinrichs & Krauwer, 2014; Piperidis, 2012)
META-SHARE	4	3	(Mapelli et al., s.d.; Piperidis, 2012)

4.2.1. Basic metadata

Studies agree on recommending these essential fields (Zervas & Sampson, 2010; Hinrichs & Krauwer, 2014; Piperidis, 2012; Park & Brenza, 2015):

- Description: (title, author, type of resource).
- Content: (language, MECR level, skills covered).
- Technical: (format, size, compatibility).
- Educational: (objectives, target groups, estimated duration).

4.2.2. Specific metadata for languages

Fourteen specific studies advocate the extension of standard metadata with linguistic fields (Hinrichs & Krauwer, 2014; Piperidis, 2012; Krajcso, 2016):

- Grammatical structure addressed.
- Thematic vocabulary (domains).
- Dialectal variety (e.g. Eastern/Western Catalan).
- Authenticity of the material (authentic/didactic).

4.2.3. Automatic metadata extraction

Eight studies (11%) propose NLP for semi-automatic metadata generation, reducing manual costs (Casali et al., 2014; Lorenzini et al., 2021). Examples:

- Extraction of nameable entities (vocabulary).
- Automatic classification of MECR level through textual complexity analysis (Lorenzini et al., 2021).
- Labelling of parts of speech for grammatical metadata (Casali et al., 2014).

4.3. Pedagogical quality criteria for resources and catalogues

There is broad consensus on the need to incorporate quality criteria that ensure consistency with learning objectives (Piedra et al., 2015; González-Blanco et al., 2017; Del et al., 2016).

Summary of NVivo results (table 3): 36/73 studies (49%) define explicit quality rubrics. The studies include proposals for automation using natural language processing techniques (Lorenzini et al., 2021).

Table 3
Dimensions of pedagogical quality (NVivo analysis of 73 studies)

Quality dimension	% of studies	(n=73)	Key indicators
Consistency of objectives	65	48	MECR alignment, clarity
Immediate feedback	42	31	Self-assessment, interactive correction

Accessibility	28	21	CC licences, WCAG 2.1
Cultural relevance	22	16	Sociocultural context, authenticity
Sustainability	18	13	Updating, maintenance

4.3.1. Consistency with learning objectives

Clarity in the structuring of content and the provision of immediate and guiding feedback are critical factors for learner autonomy (Krajcso, 2016; Law & Laurillard, 2011). Forty-eight studies (65%) emphasise the importance of explicitly specifying the expected learning outcomes (Piedra et al., 2015; González-Blanco et al., 2017; Del et al., 2016).

4.3.2. Automatic assessment of textual quality

The use of natural language processing techniques to automatically assess the accuracy of textual descriptions (Lorenzini et al., 2021) allows:

- Verify that descriptions are consistent with the actual resource.
- Detect spelling mistakes or inconsistent terminology.
- Measure the readability of instructions.

4.3.3. Repositories with quality validation

Table 4
Example of repositories with explicit quality criteria

Repository	Quality validation	Language	Source
MERLOT	Peer-review experts	Multi	Piedra et al. (2015), González-Blanco et al. (2017)
IRIS Platform	MECR validated	English	Krajcso (2016), Law & Laurillard (2011)
ORACC (Sumeri)	Scholars panel	Ancient Sumeri	Cuneiform Digital Library Initiative (n.d.)
AJOL/UPSpace	Editorial board	Multi-Africa	Anyaoku et al. (2019), Karadia & Sahoo (2021)
JClic (XTEC)	Education department	Català	Generalitat de Catalunya (2022)

4.4. Criteria related to repository architecture and resource openness

Architecture influences access to and sustainability of materials (Anyaoku et al., 2019; Hinrichs & Krauwer, 2014). The use of the OAI-PMH protocol is the standard for metadata harvesting and semantic interoperability (Piedra et al., 2015; Karadia & Sahoo, 2021).

Quantitative results (table 5): 27 studies (36%) explicitly address architecture. 22 studies (30%) recommend interoperability via OAI-PMH (Piedra et al., 2015; Karadia & Sahoo, 2021). 15 studies (20%) highlight the role of Creative Commons licences in reuse (Muñoz & Román-Mendoza, 2018; Hernández-Muñoz & Román-Mendoza, 2018; Karadia & Sahoo, 2021).

Table 5
Recommended architectural components

Architectural component	Recommendations	Implications
OAI-PMH	Mandatory implementation	Allows federation of catalogues
CC licences	CC BY, CC BY-SA by default	Facilitates educational reuse
FAIR principles	Findable, Accessible, Interoperable, Reusable	Global visibility and reuse
UX navigation	Intuitive interfaces, advanced search	Improved user experience
Digital preservation	Standard formats, backups	Long-term sustainability

4.4.1. Openness and licences

The degree of openness, through Creative Commons licences, is key to facilitating reuse and adaptation (Hernández-Muñoz & Román-Mendoza, 2018; Karadia & Sahoo, 2021). Fifteen studies (20%) advocate flexible licensing models, allowing for pedagogical modifications (Muñoz & Román-Mendoza, 2018; Hernández-Muñoz & Román-Mendoza, 2018).

4.4.2. Interoperability and OAI-PMH

Twenty-two studies (30%) indicate that the use of the OAI-PMH (Open Archives Initiative Protocol for Metadata Harvesting) protocol guarantees federation between catalogues (Piedra et al., 2015; Karadia & Sahoo, 2021). This is particularly relevant for minority languages such as Catalan, where the consolidation of scattered resources is essential (Hinrichs & Krauwer, 2014; Piperidis, 2012).

4.4.3. Interfaces and user experience

12 studies (16%) emphasise the importance of intuitive navigation and advanced search functionalities to improve the user experience (Piedra et al., 2015; Karadia & Sahoo, 2021). Key aspects:

- Multidimensional filtering (level, ability, theme).
- Facet search with visualisations.
- Personalisation of results according to profile.
- Compatibility with mobile devices.

4.5. Criteria for support in blended learning contexts

Catalogues must support situations where formal education is combined with autonomous learning (Muñoz & Román-Mendoza, 2018; Krajcso, 2016).

NVivo analysis: 31 studies (42%) explicitly address blended contexts. The emerging categories are (table 6):

Table 6
Degrees of autonomy in blended contexts (31 studies)

Level of autonomy	Characteristics	Examples of use
Low	Teacher guidance, clear structure, constant feedback	Face-to-face class with digital support
Medium	Semi-autonomous, graduated autonomy, instructions	Homework assignments, workshops
High	Self-directed, self-taught, communities	Out-of-classroom learning, MALL

4.5.1. Graduated autonomy

19 studies point to the need for criteria that inform the degree of autonomy required for each resource (Muñoz & Román-Mendoza, 2018; Krajcso, 2016). Indicators:

- *Low*: Teaching material with an explicit structure (e.g. video tutorial with questionnaire).
- *Medium*: Resources with clear instructions but without constant tutoring (e.g., self-assessed interactive exercise).
- *High*: Resources for self-directed discovery (e.g., authentic linguistic corpora, podcasts).

4.5.2. Web 2.0 and collective intelligence

15 studies (20%) highlight the potential of Web 2.0 for creating communities of practice through:

- User ratings and reviews that enrich metadata (Matthews et al., 2010).
- Folksonomies (user tags) that complement formal classification (Matthews et al., 2010).
- Collaborative forums for resolving linguistic doubts (Muñoz & Román-Mendoza, 2018).
- Design of personalised learning paths based on usage data (Agosti et al., 2011; Slimani et al., 2016).

4.5.3. Integration of collaborative tools

Ten studies propose the integration of collaborative tools (Google Docs, Padlet, etc.) to enable teachers and learners to work on shared activity objects within the catalogue or online (Agosti et al., 2011; Slimani et al., 2016).

4.6. Summary of results by thematic area

Table 4
Coverage of thematic areas in the 73 studies (NVivo analysis)

Thematic area	Studies	% (n=73)	Thematic maturity
Classification	41	55	Consolidated
Metadata	42	57	Consolidated
Pedagogical quality	36	49	Emerging
Architecture	27	36	Emergent
Mixed contexts	31	42	Emergent
Intersection (all 5 axes)	19	26	Holistic

The synthesis reveals that criteria related to classification and metadata are well established, while quality, architecture and support for mixed contexts emerge as areas requiring further development and empirical validation (table 7).

5. DISCUSSION

5.1. Contribution of the review to the literature on digital resource catalogues

The results of this systematic review confirm that, despite the existence of extensive literature on digital resources and educational technologies, there are still few studies that comprehensively address digital resource catalogues as pedagogical tools with explicit criteria for classification, description and quality (Muñoz & Román-Mendoza, 2018; Zervas & Sampson, 2013).

In this sense, one of the main contributions of this work lies in the systematisation of these criteria based on a rigorous analysis of 73 scientific studies. While many studies are limited to the technical description of repositories (Anyaoku et al., 2019), this review has identified a lack of consensus on quality standards for the textual description of metadata (Lorenzini et al., 2021).

Unlike other research that focuses on specific repositories or specific technologies, this study adopts a cross-cutting perspective that allows for the identification of recurring patterns and shared dimensions among various initiatives (Piperidis, 2012; Hinrichs & Krauwer, 2014). The organisation of the results into structured criteria contributes to a clearer conceptualisation of the role of catalogues in language learning, providing a conceptual basis that can be transferred to different educational contexts (Krajcso, 2016; Park & Brenza, 2015).

5.2. Consistency between classification, metadata and pedagogical quality

The discussion of the results shows that the criteria identified in the review do not operate in isolation, but rather form an interdependent system. The pedagogical usefulness of catalogues depends on the interrelationship between classification systems, metadata structures and pedagogical quality criteria (Krajcso, 2016; Zervas & Sampson, 2010).

Classification systems facilitate initial organisation according to linguistic skills (Bin-Hady & Al-Tamimi, 2021), but they are insufficient if they are not accompanied by metadata that allow the pedagogical characteristics to be described accurately (Park & Brenza, 2015; Di Martino, 2009). Similarly,

comprehensive metadata lose their potential if they are not guided by quality criteria that help assess the relevance of resources in relation to learning objectives (González-Blanco et al., 2017; Piedra et al., 2015). Pedagogical quality therefore emerges from the coherent integration of these elements within a holistic design (Muñoz & Román-Mendoza, 2018).

5.3. Catalogues as pedagogical tools in blended learning contexts

Digital resource catalogues acquire a particularly relevant pedagogical value when they respond to blended learning contexts, where formal education is combined with autonomous learning processes (Hernández-Muñoz & Román-Mendoza, 2018; Krajcso, 2016). In this sense, criteria related to the degree of autonomy required, the clarity of instructions and the possibility of use outside the classroom emerge as key elements (Kárpáti, 2009; Zervas & Sampson, 2010).

From this perspective, catalogues can act as tools for pedagogical mediation, facilitating the informed selection of resources consistent with curricular objectives (Muñoz & Román-Mendoza, 2018).

This mediation function is enhanced when catalogues incorporate personalised recommendation systems (Slimani et al., 2016), which tailor the search experience to the individual profiles of learners and teachers. Federated repository architectures (CLARIN, META-SHARE) exemplify this evolution towards open and interoperable systems that facilitate the visibility of quality linguistic resources (Hinrichs & Krauwer, 2014; Piperidis, 2012).

5.4. Implications for languages with non-hegemonic digital ecosystems: the case of Catalan

The interpretation of the results takes on special relevance in the case of languages with non-hegemonic digital ecosystems. Despite having a significant set of resources, the literature highlights the lack of clear pedagogical systematisation in the organisation of these materials (Kárpáti, 2009; Mohamad et al., 2011). In the case of Catalan, the application of these criteria could guide the design of catalogues that maximise the impact of existing resources, ensuring interoperability and accessibility through quality metadata (Moreno et al., 2019; Piedra et al., 2015).

The recommendations systematised in this review are particularly relevant for initiatives such as the expansion of XTEC (Xarxa Telemàtica Educativa de Catalunya) or the creation of university catalogues dedicated to Catalan. The adoption of criteria based on OAI-PMH standards, open CC licences and IEEE LOM metadata would ensure sustainability and integration with European infrastructures such as CLARIN (Hinrichs & Krauwer, 2014; Piperidis, 2012; Piedra et al., 2015).

Furthermore, collaboration with minority language initiatives (ELRA, European Language Resources Association) could facilitate access to resources and best practices that have proven effective in similar contexts (Piperidis, 2012; Mapelli et al., 2018).

5.5. Limitations of the study and future lines of research

Despite its contributions, some limitations must be acknowledged. The review was based on international databases (Scopus and Web of Science), which may have excluded relevant studies in other languages or local sources (Page et al., 2021). Furthermore, the study focuses on the analysis of

scientific literature and does not yet include empirical validation with end users (teachers or learners) (González-Blanco et al., 2017; Del et al., 2016).

Future lines of research include the application of systematised criteria to the design and evaluation of real catalogues and the analysis of their impact on teaching practices (González-Blanco et al., 2017). It would also be relevant to:

1. Conduct case studies of minority language syllabuses (e.g. XTEC-Català).
2. Develop prototypes of syllabuses that implement the identified criteria.
3. Validate the criteria with groups of Catalan teachers and learners.
4. Study the long-term sustainability of federated repositories.

6. CONCLUSIONS

This study has identified and systematised criteria for the design and use of digital resource catalogues for language learning through the qualitative analysis of 73 studies. The research concludes that:

1. Educational value: Catalogues acquire real value when they incorporate clear classification criteria (MECR skills and levels), metadata adapted to linguistic specificities, and explicit pedagogical quality criteria (Muñoz & Román-Mendoza, 2018; Krajcso, 2016).
2. Architecture and Openness: The accessibility, reusability, and sustainability of materials depend directly on the architecture of repositories and the use of open licences (Creative Commons), especially in interoperability protocols such as OAI-PMH (Anyaoku et al., 2019; Karadia & Sahoo, 2021).
3. Mixed Contexts: In blended learning, continuity between the classroom and personal space requires flexible tools that clearly inform users of the degree of autonomy necessary for the use of resources (Hernández-Muñoz & Román-Mendoza, 2018; Kárpáti, 2009).
4. Applicability: The systematisation presented provides a sound basis for improving language catalogues with non-hegemonic digital ecosystems, where pedagogical organisation often does not follow the availability of resources (Mohamad et al., 2011).

Finally, this work lays the foundations for future lines of research aimed at the empirical validation of the criteria and the analysis of their real impact on language learning processes in diverse contexts.

The results obtained show that digital resource catalogues acquire educational value when they incorporate clear classification criteria, metadata systems appropriate to the specificities of language learning, and explicit pedagogical quality criteria. The coherent integration of these elements facilitates the informed selection of resources, supports teaching decisions and promotes autonomous learning.

Furthermore, the study highlights the importance of considering the architecture of repositories and the degree of openness of resources as key factors in ensuring the accessibility, reusability and sustainability of digital materials. These aspects become particularly significant in blended learning contexts, where continuity between formal education and learning outside the classroom requires flexible and adaptable tools.

From an applied perspective, the systematisation of criteria presented offers a sound basis for guiding the design and improvement of digital resource catalogues for language learning. In particular, the results are especially relevant for languages with non-hegemonic digital ecosystems, such as Catalan,

where the availability of resources is not always accompanied by a pedagogical organisation that facilitates their educational use.

Finally, this work lays the foundations for future lines of research aimed at the application and empirical validation of the criteria identified, especially in contexts of language planning and national education policies.

7. REFERENCES

Agosti, M., Ferro, N., & Silvello, G. (2011). *Semantic information object retrieval. Foundations and theory*. Springer.

Anyaoku, E. N., Echedom, A. U. N., & Baro, E. E. (2019). Digital preservation practices in university libraries: An investigation of institutional repositories in Africa. *Digital Library Perspectives, 35*(1), 41–64. https://doi.org/10.1108/DLP-10-2017-0041

Baldwin, T., Bird, S., & Hughes, B. (2006). *Collecting low-density language materials on the web*. In Proceedings of the 12th Australasian Web Conference (pp. 1–8).

Bin-Hady, W. R. A., & Al-Tamimi, N. O. M. (2021). The use of technology in informal English language learning: Evidence from Yemeni undergraduate students. *Learning and Teaching in Higher Education: Gulf Perspectives, 17*(2), 107–120. https://doi.org/10.1108/LTHE-09-2020-0037

Bourda, Y., & Doan, B. (2003). *The semantic Web for learning resources*. In Proceedings 3rd IEEE International Conference on Advanced Technologies (pp. 1–8). https://doi.org/10.1109/ICALT.2003.1215104

Burke, M., & Zavalina, O. L. (2020). Identifying challenges for information organisation in language archives: Preliminary findings. In *Sustainable digital communities* (pp. 622–629). Springer. https://doi.org/10.1007/978-3-030-43687-2_52

Casali, A., Deco, C., & Beltramone, S. (2014). *Automatic gathering of educational digital resources to populate repositories.* In Proceedings of the XV International Conference on Human Computer Interaction (pp. 1–8). https://doi.org/10.1145/2662253.2662339

Churchill, N. (2017). Digital media and learning: Transformative educational approaches. In *Educational technology and learning* (pp. 1–25). Springer.

Cuneiform Digital Library Initiative (CDLI). (n.d.). *ORACC: Open richly annotated cuneiform corpus*. Retrieved from http://oracc.museum.upenn.edu/

Del, S. R., Spivack, J., Barrera-Marín, S., & Fiallos, F. (2016). Quality criteria for digital educational resources. *International Journal of Educational Technology in Higher Education, 13*(1), 1–15.

Department of Education of the Government of Catalonia. (2022). *Plan for the promotion of the Catalan language in the digital and technological fields*.

Di Martino, J. (2009). Classification and categorisation in digital learning environments. *Journal of Digital Learning, 5*(2), 45–63.

European Language Resources Association. (2024). *Language policy report 2022*. Sociolinguistic perspectives on language planning.

Generalitat de Catalunya, Department of Education. (2022). *XTEC: Educational telematic network of Catalonia.* Digital educational resources platform.

Generalitat de Catalunya, Department of Culture. (2022). *An egalitarian sociolinguistic framework for the Catalan language.* https://llengua.gencat.cat

González-Blanco, E., Seoane-Cegarra, R., & Riofrío-Luzcando, G. (2017). Digital humanities and digital education: A review of practices and pedagogies. *Digital Humanities Quarterly, 11*(3), 1–20.

Gonzalez, J. L., Van't Hooft, A., Carretero, J., & Sosa-Sosa, V. J. (2017). Nenek: A cloud-based collaboration platform for the management of Amerindian language resources. *Language Resources and Evaluation, 51*(4), 897–925. https://doi.org/10.1007/s10579-016-9361-8

Hernández-Muñoz, N., & Román-Mendoza, E. (2018). Learn with me: Requirements of the digital age for good practices in second language teaching. *Circle of Applied Linguistics in Communication*, 76, 31–48. https://doi.org/10.5209/CLAC.62496

Hinrichs, E., & Krauwer, S. (2014). *The CLARIN research infrastructure: Resources and tools for eHumanities scholars*. In Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14) (pp. 1525–1531). Reykjavik, Iceland: European Language Resources Association (ELRA).

Kárpáti, A. (2009). *Web 2 technologies for Net Native language learners: A "social CALL*". ReCALL, 21(2), 139–156. https://doi.org/10.1017/S0958344009000160

Karadia, A., & Sahoo, D. J. (2021). A comparative study of India and Australia open access repositories in OpenDOAR. *IP Indian Journal of Library Science and Information Technology, 6*(1), 57–60. https://doi.org/10.18231/j.ijlsit.2021.013

Kieu, T. A., Phung, N. H., Long, N. N., & Trang, H. D. (2021). The effectiveness of digital resources in English language learning: A systematic review. *Journal of Language Teaching and Research, 12*(5), 234–248.

Krajcso, Z. (2016). Classification and quality criteria for Open Educational Resources in the field of foreign language learning. *Journal of Language and Cultural Education, 4*(1), 48–59. https://doi.org/10.1515/jolace-2016-0004

Law, N., & Laurillard, D. (2011). Learning design as a medium for scaffolding teacher learning and collaboration. In *Digital technologies and learning* (pp. 1–20). Springer.

Lorenzini, M., Rospocher, M., & Tonelli, S. (2021). Automatically evaluating the quality of textual descriptions in cultural heritage records. *International Journal on Digital Libraries, 22*(2), 217–231. https://doi.org/10.1007/s00799-021-00302-1

Mapelli, V., Arranz, V., Mazo, H., Kamocki, P., & Popescu, V. (2018). *New directions in ELRA activities*. In Proceedings of the 11th Language Resources and Evaluation Conference (LREC 2018) (pp. 1–6). European Language Resources Association (ELRA).

Matthews, B., Jones, C., Puzoń, B., Moon, J., Tudhope, D., Golub, K., & Lykke Nielsen, M. (2010). *An evaluation of enhancing social tagging with a knowledge organisation syste*m. Aslib Proceedings, 62(4/5), 447–465. https://doi.org/10.1108/00012531011074690

Mohamad, N. A. A., Hassan, T. F. N. T., Muda, T. N. T., Aziz, N. A., & Ishraf, A. H. (2011*). Multilingual online resources for minority languages of a campus communit*y. Procedia - Social and Behavioural Sciences, 27, 291–298. https://doi.org/10.1016/j.sbspro.2011.10.610

Moreno, A., Torre, D., Valverde, A., & Campillos, L. (2019). Study on reusable documents as linguistic resources within the framework of the Language Technology Promotion Plan. *Natural Language Processing*, 63, 167–170.

Muñoz, H., & Román-Mendoza, E. (2018). Learn with me: Requirements of the digital age for good practices in second language teaching. *Circle of Applied Linguistics in Communication*, 76, 31–48. https://doi.org/10.5209/CLAC.62496

Page, M. J., McKenzie, J. E., Bossuyt, P. M., et al. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *The BMJ, 372*(71).

Park, J., & Brenza, A. (2015). Evaluation of semi-automatic metadata generation tools: A survey of the current state of the art. *Information Technology and Libraries, 34*(3), 22–42. https://doi.org/10.6017/ital.v34i3.5889

Piedra, N., López, M. J., Tovar, E., Colomo-Palacios, R., & Chicaiza, J. (2015). Framework for the integration of digital resources based on a Semantic Web approach. *RISTI - Revista Ibérica de Sistemas e Tecnologias de Informação, 3,* 55–70. https://doi.org/10.17013/risti.e3.55-70

Piperidis, S. (2012). The META-SHARE language resources sharing infrastructure: Principles, challenges, solutions. In N. Calzolari et al. (Eds.), *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12) (*pp. 36–42). Istanbul: European Language Resources Association (ELRA).

Reinders, H. (2017). Digital games and second language learning. In *Language, education and technology* (pp. 1–20). Springer. https://doi.org/10.1007/978-3-319-02237-6_26

Slimani, H., El Faddouli, N., Benslimane, R., & Bennani, S. (2016). *Personalised search and recommendation in a digital educational resources repository: The case of ORI-OAI*. In 2016 4th IEEE International Colloquium on Information Science and Technology (CiSt) (pp. 1–6). https://doi.org/10.1109/CIST.2016.7805108

Tobar, C. M., & De Lima, S. I. (2007). *Assessing a learning object standard adherence*. In 2007 37th Annual Frontiers in Education Conference - Global Engineering: Knowledge Without Borders, Opportunities Without Passports. https://doi.org/10.1109/FIE.2007.4418177

Winslow, S. M., Schneider, G., Bleier, R., Steiner, C., Pollin, C., & Vogeler, G. (2019). *Ontologies in the digital repository: Metadata integration, knowledge management and ontology-driven applications*. In Proceedings of WODHSA 2019: 1st International Workshop on Ontologies for Digital Humanities and their Social Analysis (CEUR Workshop Proceedings, Vol. 2518, pp. 1–12). IAOA.

Yan, Z., & Xiao-yue, G. (2007). Knowledge maps and semantic search for digital library resources. *Journal of Information Science, 33*(5), 567–581.

Zervas, P., & Sampson, D. G. (2010). *Enhancing educational metadata with Mobile Assisted Language Learning information*. In 2010 10th IEEE International Conference on Advanced Learning Technologies (pp. 1–8). https://doi.org/10.1109/ICALT.2010.79

Zervas, P., & Sampson, D. G. (2013). A quantitative analysis of the reuse of Mobile Assisted Language Learning resources: The case of Mobile2Learn repository. In *2013 IEEE 13th International Conference on Advanced Learning Technologies* (pp. 1–8). https://doi.org/10.1109/ICALT.2013.16

Zervas, P., Fragkos, K., & Sampson, D. G. (2013*). ASK LDT 2.0: A web-based graphical tool for authoring learning designs*.

INFORMATION ABOUT THE AUTHORS

Mònica Vilosa Sánchez
Departament d'Educació i Formació Professional, Generalitat de Catalunya

Licenciada en Filología Catalana. Profesora de Educación Secundaria y Bachillerato. Directora de la escuela de adultos CFA Girona Nou. Doctoranda del Programa de Doctorado en Educación de la Universitat de Girona.

Juan González-Martínez
Departament de Pedagogia. Universitat de Girona

Profesor Agregado Serra Húnter. Doctor en Lingüística Aplicada (UNED) y Tecnología Educativa (URV). Licenciado en Antropología Social y Cultural (UNED). Miembro del grupo de Investigación UdiGital.Edu de la UdG y profesor del Doctorat en Educació de la UdG y del Doctorado en Educación Multimodal de la Universidad Autónoma de Querétaro. Coordinador del Grado de Maestro/a en Educación Primaria.