



HAL
open science

Modelling of a Heterogeneous Corpus: The Example of Chapbook Literature

Elina Leblanc

► **To cite this version:**

Elina Leblanc. Modelling of a Heterogeneous Corpus: The Example of Chapbook Literature. Digital Studies / Le champ numérique, 2023, 13 (1), 10.16995/dscn.8091 . hal-04271423

HAL Id: hal-04271423

<https://hal.science/hal-04271423>

Submitted on 6 Nov 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License



Open Library of Humanities

Modelling of a Heterogeneous Corpus: The Example of Chapbook Literature

Elina Leblanc, University of Geneva, Switzerland, Elina.Leblanc@unige.ch

This article proposes an analysis of Spanish chapbook literature from the digital perspective. It provides a systematic study of the metadata and services selected by seventeen digital libraries to model their collections.

In the first part, we propose an overview of those libraries. Their great variety reflects the heterogeneity of this popular literature, which is at the margin of studies on printed productions, and escapes the classifications traditionally used for books. In the second part, we describe four different strategies used by digital libraries to model this type of content, focusing either 1) on the document as an archive to be preserved, 2) on the document as the result of an editorial process, 3) on the text, or 4) on the illustrations. Our objective is thus to outline a digital model for these types of documents and to help future projects in defining their own offer of services.

This article is part of the research project "Untangling the cordel / Démêler le cordel / Desenrollando el cordel (2020–2024)" financed by the Philanthropic Sandoz-Monique de Meuron Family Foundation and directed by the professor Constance Carta (University of Geneva).

Cet article propose une analyse de la littérature de colportage espagnole au prisme du numérique, en étudiant de manière systématique les métadonnées et les services sélectionnés par dix-sept bibliothèques numériques pour modéliser leurs collections.

Dans un premier temps, nous proposons un panorama de ces bibliothèques, dont la grande variété reflète celle de cette littérature populaire, à la marge des études sur les productions imprimées et qui échappe aux classifications traditionnellement employées pour le livre. Dans un second temps, nous décrivons quatre stratégies différentes employées par les bibliothèques numériques pour modéliser ce type de contenus, en mettant l'accent soit sur le document en tant qu'archives à conserver, soit sur le document en tant que résultat d'un processus éditorial, soit sur le texte, soit sur l'illustration. Notre objectif est ainsi d'esquisser les contours d'un modèle numérique pour ce type de documents et d'aider de futurs projets dans la définition de leur propre offre de services.

Cet article s'inscrit dans le cadre du projet de recherche « Démêler le cordel / Desenrollando el cordel / Untangling the cordel (2020–2024) » financé par la Fondation philanthropique Famille Sandoz-Monique de Meuron et dirigé par la professeure Constance Carta (Université de Genève).



Introduction

Chapbook literature refers to heterogeneous, often short, inexpensive, and mass-printed productions, published since the beginning of printing (Gomis and Botrel 2019). In Spain, these productions can take the form of *pliegos sueltos* or *pliegos de cordel* due to their shape, a sheet folded several times, and their mode of sale, being hung with clothespins on strings (Botrel 2000; Gomis and Botrel 2019). Composed of a few pages, printed on low-quality paper, and illustrated with woodcuts, *pliegos de cordel* appear as a popular literature for the less educated people (see **Figure 1**). It is “a ‘frontier’ genre, with real ‘classics of the people,’ made of adaptations and appropriations” of scholarly literature (Botrel 2000). In verse or prose, both written and oral, they are characterized by a strong disparity in the types of text and themes addressed, such as narratives of events, which prefigure the periodic press, fiction, religious writings, songs, poems, plays, calendars, etc. (Gomis and Botrel 2019). Despite this great diversity of genres and themes, which makes difficult any work of taxonomy and modelling (Gomis and Botrel 2019) one constant remains: their mode of production. Indeed, chapbooks share a common layout and typography that gives them a visual identity. The *literatura de cordel* (or chapbook literature) can thus be considered as a publishing genre with its own codes and strategies (Botrel 2001).



Figure 1: Example of a pliego de cordel (*Trobos discretos y divertidos, para entretener a los curiosos* 1858).

Numerous digital libraries dedicated to *literatura de cordel* have emerged since the 1990s, as evidenced by the Catálogo y Biblioteca Digital de Relaciones de Sucesos (CBDRS), which has offered, since 1993, a digital portal for narratives of events produced between the 16th and 18th centuries (Pena Sueiro and García 2014). These digital projects are the result of large-scale initiatives of the collecting, cataloguing, and digitizing of printed materials held in many public and private institutions.

Just as *pliegos* are characterized by the heterogeneity of their content, digital libraries that disseminate them are also different in nature. They reflect the pluridisciplinary aspect of this literature, which interests many research communities (book history, literature, history, musicology, art history, etc.). The challenge for these libraries is to model the *pliego* as a hybrid object at the frontier between document, text, image, and speech, and to manage this complexity within their interface.

This article proposes an analysis of the strategies and approaches adopted by digital libraries to model and disseminate chapbook literature through a systematic study of their metadata and the resulting services. Although we are primarily interested in Hispanic chapbooks, to broaden our focus, here we have chosen to include digital projects dedicated to international equivalents of Hispanic chapbooks that share comparable features, such as Anglo-Saxon broadsides, French *complaintes*, or Brazilian *folhetos*.

Through this state of the art, the objective is first to highlight the contributions of digital technologies for the study of these “non-books,” which challenge the models traditionally used for the digital representation of books. We also identify the common points between these different libraries. We thus want to outline digital models for this type of document and to assist future projects in defining their own service offers.

Where to find chapbook literature online? A typology of digital libraries

We classified the seventeen projects included in our study into three types of digital libraries: *general libraries*, where chapbooks are a sub-collection of a larger collection; *specialized portals* that aggregate collections of several partner institutions; and *thematic libraries* dedicated to a particular collection or genre.

General libraries (see **Table 1**) are, in our case, digital counterparts of a physical library (the Biblioteca Nacional de Chile, the Cambridge University Library), or of heritage and academic institutions (the Universidad de Chile, Fundação Casa de Rui Barbosa, Generalitat de Catalunya). Within these libraries, chapbook literature appears as a sub-collection among a larger group of contents. The challenge for these general libraries is to adapt to all types of content, regardless of their theme, geographical origin, or time period. They target a varied public through a sober interface and services dedicated to bibliographical research and simple consultation. They resemble data providers whose aim is to offer an efficient access to their content (Mion-Mouton 2012).

Project's name	Institutions	Type of content	Point of view
Cambridge University Digital Library (CUDL)	Cambridge University	4500 <i>pliegos</i> (18 th -19 th centuries)	Librarian Textual
Repositório Rui Barbosa de Informações Culturais (RUBI)	Fundação Casa de Rui Barbosa (FCRB - Brazil)	9000 sheets (from the 1960s)	Librarian
Coleccion Literatura oral y Tradiciones populares	Biblioteca National Digital de Chile	327 <i>pliegos</i> (19 th -20 th centuries) ; 288 <i>folletos</i> (19 th -20 th centuries)	Librarian
Coleccion Lira Popular	Universidad de Chile	860 <i>hojas volantes</i>	Librarian
Calaix	Generalitat de Catalunya	3814 <i>romanços</i>	Librarian

Table 1: Description of the general libraries.

Specialized digital portals (see **Table 2**) function as aggregators that build fictitious collections from the contents of various institutions on the model of Europeana or the Digital Public Library Americana (DPLA). However, they differ from these generalist aggregators by offering collections-focused content that is typologically homogenous. For example, the portal Mapping Pliegos is dedicated to 19th-century Spanish *pliegos de cordel*; the CBDRS project presents narratives of events (*relaciones de sucesos*), a specific type of *pliego*; the portal Comedias Sueltas USA offers Spanish plays published in the form of sheets (*suestras*) and preserved in North American libraries. These portals provide users with a single point of access that overcomes the typological fragmentation of collections in multiple institutions. They do not offer consultation of digitized chapbooks *via* their interface, but through the interface of the source digital libraries. Only the metadata necessary to identify the documents are directly accessible. These portals do not replace the other libraries from which they draw their content but are complementary tools that help users target their research. Their services are thus dedicated to bibliographical research. However, unlike general libraries, these services go beyond simple and advanced searches. They provide additional functionalities linked to the nature of their collections with, for instance, the implementation of indexes or thematic maps (Leblanc 2019).

Finally, *specialized digital libraries* (see **Table 3**) are designed for a particular collection or theme, such as the Biblioteca Virtual Cordel, which is dedicated to the Brazilian *folhetos* collection of the University of Poitiers, or the McGill Library's Chapbook Collection, which publishes its English and American chapbooks (17th-19th centuries). They are characterized by a strong contextualization of their data, allowed by their narrower field of study. Moreover, some of them are the expression of research work on chapbook literature and propose a reading of contents from this specific point of view (Mion-Mouton 2012).

Project's name	Institutions	Type of content	Point of view
Catálogo y biblioteca Digital de Relaciones de Sucesos (CBDRS)	Universidad de Coruña (Spain)	5958 narratives of events	Librarian Editorial
Mapping Pliegos	Collective	25,000 <i>pliegos</i> from the 19 th century	Librarian Editorial Textual Iconographic
Comedias Sueltas USA	Alliance of several American libraries and academic institutions	4615 <i>sueitas</i> , the 17 th century and the beginning of the 19 th century	Librarian Editorial Iconographic
Literatura de cordel y teatro en España	Project of Santiago Cortés Hernández (Doctoral thesis)	450 <i>pliegos</i> between 1675 and 1825	Textual
Complaintes Criminelles	<i>CriminoCorpus</i> (Museum of Justice, Crimes and Sentences - France)	1157 <i>complaintes</i> (between 1870 and 1939)	Librarian Editorial

Table 2: Description of specialized digital portals.

Project's name	Institutions	Type of content	Point of view
Fundación Joaquín Díaz - Biblioteca	Fundación Joaquín Díaz (Spain)	6675 <i>pliegos</i>	Librarian Editorial Textual
Biblioteca Virtual Cordel	University of Poitiers	4000 <i>fohletos</i>	Librarian Editorial Textual
John Johnson Collection	Bodleian Library	67,000 ephemera	Librarian Editorial Iconographic
Impresos Populares Iberoamericanos (IPI)	Collective (but the website is carried by the UNAM - Mexico)	876 <i>cuadernillos</i> , 1818 <i>hojas volantes</i> , 114 <i>pliegos de cordel</i> , 114 <i>librillos</i> , 6 <i>libros</i>	Librarian Editorial Textual
McGill Library's Chapbook Collection	McGill Library (Canada)	900 <i>chapbooks</i>	Librarian Editorial Textual
The Word on the Street	National Library of Scotland	1800 <i>broadsides</i> between 1650 and 1910	Editorial Textual
Broadside Ballads Online	Bodleian Library and Oxford University	34,000 <i>broadsides</i>	Librarian Editorial Iconographic

Table 3: Description of the specialized digital libraries.

“Tell me your metadata and I’ll tell you who you are”: Metadata as a mirror of strategies for representing chapbook literature

Metadata, whether descriptive or technical, “allow us to move from a series of files to something that we can recognize as a document” (Bermès 2011, 295). As veritable “digital binders” (Bermès 2011, 292), they are necessary for the management of any collection, whether for conservation, classification, research, or navigation purposes (Bermès 2011, 295).

The choice of metadata is not trivial since it is the cornerstone of all digital libraries. It defines their organization, as well as their services. Whether it is the creation of search forms, maps, and indexes, or the proposal of paths in the collections, all these services are based on the metadata that describe the contents. Metadata thus contribute to the process of editorialization of content and knowledge, as defined by Marcello Vitali-Rosati. They shape and influence the way a user perceives and evaluates the information offered by the library. They create scientific contexts and offer various interpretations and points of view of a content, depending on the services with which they are presented (Vitali-Rosati 2016). Choosing metadata is both giving an orientation to one’s project (librarian, literary, historical, ethnographic, etc.), and defining the appearance of one’s library. Therefore, their analysis allows us to learn more about the approaches adopted by the digital libraries of our study to model and present their collections of chapbooks to their users.

1. Overview of the metadata chosen by the digital libraries dedicated to chapbooks

The metadata most present in the libraries studied are the title, the date, the name of the printer, the place of publication, the format of the document, and thematic keywords. Next are metadata about the author, the condition of the document, illustrations, and the holding institution, which is a central piece of information for digital portals. The rest of the metadata varies greatly between projects: for example, if the mention of the incipit seems common to specialized digital libraries, the presence of the name of the holding institution is more common to portals.

The choice of metadata reflects the type of digital library. Let’s take an example with the Digital National Library of Chile and its collection about *Literatura oral y Tradiciones populares* (see **Figure 2**).

The information presented here goes to the essential. It is necessary to help users identify a document and distinguish it from other prints that might have a similar title or layout. This strategy is specific to generalist digital libraries. The choice of a simple and generic metadata model can be applied to their whole collection, regardless of the nature or the time period of the documents. Thus, in the example in **Figure 3**, we have randomly selected a content in the same library as in the previous example.

Clasificaciones

Autor Secundario: [Becerra M., F.](#)

Colección: [Archivo de Literatura Oral y Tradiciones Populares / Folletos de poesía popular y cancioneros](#)

Materias: [Música para arpa - Romances \(Música\) - Valses - Canciones folklóricas - Canciones para brindar - Cuecas - Himnos nacionales - Tonadas - Música para guitarra y piano - Habaneras](#)

Tipo de objeto: [Libro](#)

Año: [1908](#)

Datos de publicación: Santiago de Chile : Imprenta i Encuadernación Universitaria, 1908.

Tipo de acceso: Acceso en línea

BN Código: LO0013031

N° Sistema: 1085540

BND id: 616974

Figure 2: Extract of the metadata of the Digital national library of Chile (*La Polola, o, La reina de mis cantos* 1908).

Clasificaciones

Autor: [Rollin, M. Charles, 1661-1741](#)

Colección: [Fondo General / Colección General](#)

Materias: [Historia - Emperadores romanos - Roma](#)

Tipo de objeto: [Libro](#)

Año: [1795](#)

Datos de publicación: Madrid : En la Oficina de Don Blas Roman, 1795-1797.

Tipo de acceso: Acceso en línea

BN Código: FG0000026

N° Sistema: 52328

BND id: 618554

Figure 3: Extract of the metadata of the Digital national library of Chile (Rollin 1795).

We notice that the information given is identical, although this is a Roman history book from the late 18th century, and not a chapbook. The same observation can be made for the five generalist libraries of our corpus in which the metadata are rather generic to cover collections of varied documents (see **Table 4**).

Digital library	Metadata
Cambridge University Digital Library (CUDL)	Title, holding institution, classmark, thematic keywords, printer, publication place, date, physical description, format, material
Repositório Rui Barbosa de Informações Culturais (RUBI)	Title, author, date, publication place, thematic keywords
Coleccion Literatura oral y Tradiciones populares	Title, author, collection's name, document type, text type, data, digital access, digital identifiers, format, physical description, table of contents
Coleccion Lira Popular	Title, author, language, thematic keywords, format, copyrights, information about digitization
Calaix	Title, URI, format, date, digital collection, abstract, physical description, thematic keywords, language, holding institution, copyright, table of contents

Table 4: List of metadata chosen by generalist libraries.

The choice of a generic model to describe chapbooks can be explained in several ways. For the librarian, it is a means to facilitate the cataloguing. It does not imply a wide knowledge of the whole collection of a library, as the evolutions of the profession prevent having an encyclopedic expertise on all the documents (Leblanc 2019). For users, it ensures a unified interface that simplifies searches and navigation through the digital library. Thus, for these types of projects, metadata don't depend on the nature of data. Rather, they are related to institutional and librarian imperatives.

Specialized portals and thematic libraries, on the contrary, offer descriptions linked to the nature of their collections. As they are focused on a specific type of chapbooks or a specific time period, these projects have the possibility to define a more precise and elaborate metadata model. The portal CBDRS adopts a bibliographical approach (typography, collation of signatures, catchwords, initials, etc.). Similarly, the Impresos Populares Iberoamericanos (IPI) library gives information about the context of production and sale of 19th-century chapbooks (see **Table 5**). The level of granularity is more accurate and highlights different ambitions between general libraries and specialized ones. The latter appear as laboratories of research and experiment. Therefore, the choice of metadata is linked not only to librarian criteria, but also to research considerations. They depend on the institutional and scientific context in which they have been conceived.

Digital library	Metadata
Catálogo y Biblioteca Digital de Relaciones de Sucesos (CBDRS)	Title, type of document, date, digital identifier, publication place, printer, format, number of pages, signature, typography, presence of catchwords, initials, license, bindings, type of discourse, type of writing, classmark, holding institution
Mapping Pliegos	Title, author, publication place, editor, printer, date, page number, presence of illustrations, dimensions, incipit, thematic keywords, holding institution
Comedias Sueltas USA	Holding institution, classmark, author, title, subtitle, date, imprint, language, format, signature, thematic keywords, digital identifier
Fundación Joaquín Díaz - Biblioteca	Title, incipit, author, date, printer, publication place, list of contents, list of woodcuts, commentaries
Biblioteca Virtual Cordel	Title, author, collection, type of document, language, date, page number, printer, presence of illustrations
Complaintes Criminelles	Title, language, author, editor, classmark, tone, number of verses, presence of a chorus, date, commentaries, state of conservation, summary of the text, information about the tone
John Johnson Collection	Title, digital identifiers, type of material, author, other contributors, incipit, date, physical form and production, thematic keywords, holding institution, classmark
Impresos Populares Iberoamericanos (IPI)	Title, list of contents, date, page number, type of document, dimension, type and colour of the paper, state of conservation, printer, publication place, imprint, sale place, price, digital collection, information about the digitization
McGill Library's Chapbook Collection	Title, printer, imprint, date, document description, format, language, thematic keywords, type of text, digital collection, classmark, copyright, digital identifier
The Word on the Street	Commentaries about the text, date, classmark
Literatura de cordel y teatro en España	Imprint, date, title, page number, format, catchword, column number, description of illustrations, type of text, incipit
Broadside Ballads Online	Title, incipit, printer, publication place, date, imprint

Table 5: List of metadata chosen by specialized portals and libraries.

While the metadata adopted by these seventeen projects share some common features, such as the recurrent presence of imprint information, it is difficult to find a true consensus on the best method to describe these types of prints. It illustrates the wide variety of chapbooks and the numerous scientific communities interested in this subject, such as bibliographical historians (CBDRS, IPI, Comedias Sueltas USA), literary researchers (Literatura de cordel y teatro en España), specialists of folklore and orality (Complaintes criminelles, Fundación Joaquín Díaz) or even historians (The Word on the Street).

Therefore, the metadata here better reflect the perspective each project takes on chapbooks than the library type itself. These projects make a clear distinction between document, text, and illustration. Here, by *document* we mean “physical objects that contain some sort of inscribed information” (Pierazzo 2015, 47). It consists of a set of features that only take on their meaning when they are exploited by an agent (curator, cataloguer, reader, etc.). Without this action of the agent, the document remains an “inert object” (Pierazzo 2015, 49). When the inscriptions of a document are analyzed by a reader, they become a *text* that can be invested with different meanings depending on the interests of the reader (Pierazzo 2015). In the same way, an *illustration* is also a feature of a document, a graphic inscription that can be interpreted in different ways by an agent.

Considering these definitions, we identify four strategies for modelling and representing chapbooks:

A librarian-centred approach to document preservation with a focus on bibliographical information

An approach centred on book history where the document is perceived as the result of an editorial process

An approach centred on texts

An approach centred on illustrations

These different approaches are not exclusive, but complementary. The same project can offer services that emphasize some above others to address various scientific communities. For example, in the case of *IPI*, a space dedicated to the study of chapbooks from the perspective of book history (*Búsqueda bibliográficas*) and another one dedicated to the study of texts (*Búsqueda literaria*) coexist in the same interface. The Broadside Ballads Online project also offers different types of interfaces: one that displays a catalogue of printed materials that can be filtered by categories related to the editorial process, and the other one to the search of illustrations extracted from the chapbooks.

2. Chapbooks as documents: Between preservation of an archive and analysis of an editorial process

The interest in chapbooks as documents can lead to two different modelling strategies. If the agent is a librarian, printed materials are perceived from a librarian perspective. The objective is to preserve and make chapbooks available to as wide a public as possible (Repositório Rui Barbosa de Informações Culturais, Colección Literatura oral y Tradiciones populares, Colección Lira Popular, Cambridge University Digital Library, Calaix). These institutions describe their collections with information about the printer, the date, the place of edition, the type of documents, or material aspects (format,

size, number of pages, document condition). When digital libraries preserve several types of chapbooks, the mention of the printer or the type of document is essential to organize contents for which the presence of a date or an author is inconsistent (Nieto 2015). It also allows one to distinguish them when their titles are identical or similar. For example, the Mapping Pliegos portal contains eight printings with identical title spellings (*Décimas glosadas para cantar los aficionados*), but two come from Córdoba and the other six from Madrid, with different dates and printers.

To structure their metadata, generalist libraries employ standards specific to the cataloguing of books in libraries (Nieto 2015). Among the standards used, we can mention MARC-XML, MARC 21, ISBD, or Dublin Core. However, these standards are mainly designed for cataloguing books. When they are applied to chapbooks, also called “non-books,” several fields are left empty. They also tend to minimize the importance of text or illustration, two elements that are nonetheless essential in the classification of these documents (Lambert 2015a; Lambert 2015b). Some libraries have then chosen to create their own standard, such as the John Johnson Collection at the Bodleian Library (Lambert 2015a).

It can be noted that the use of digital standards for the encoding of descriptive metadata is also an element that distinguishes generalist digital libraries from the specific ones. As we saw earlier, the former rely on librarian metadata standards, whereas the latter resort more often to “home-made” metadata, exception made for the McGill Library’s Chapbook Collection, which uses the XML-TEI standard. This tendency observed in thematic libraries and the lack of transparency about the standards used is a risk for interoperability and long-term preservation of data. This situation may be explained by the unsuitability of librarian standards to these types of documents. However, other standards like XML-TEI or XML-EAD could be used to describe chapbooks, as they offer a wider level of granularity to describe texts and documents that can adapt to the nature of chapbooks, as demonstrated by the McGill Library.

Each library in this study wishes to report, catalogue, disseminate, and reunite a disparate heritage. Often stored in boxes, these documents are invisible to users in physical library catalogues where only the boxes’ classification numbers are reported (Lambert 2015a). Digital technologies thus give them visibility and increase their accessibility.

However, not all the studied libraries are involved only in the preservation, cataloguing, and dissemination of a document. Depending on the way digital libraries use metadata, their focus may switch from a logic of preservation-dissemination to an approach centred on the analysis of chapbooks from the point of view of their production and sale (i.e., their editorial process). This is done through services that target this

specific aspect of chapbooks. We can mention again the *búsqueda bibliográfica* of the IPI library, which consists of several search forms dedicated to the printer, his address, the selling price, or the colophon (see **Figure 4**).

Consultas bibliográficas

El registro de los impresos en la base de datos guarda criterios bibliográficos, los siguientes permiten realizar consultas precisas sobre el total del inventario:

- Consulta por texto presente en el **incipit**
- Consulta por texto presente en la **tabla de contenidos**
- Consulta por el **editor**
- Consulta por la **colección o serie**
- Consulta por **dirección de imprenta o despacho**
- Consulta por el **pie de imprenta**
- Consulta por el **estado de conservación**
- Consulta por el **precio** de venta

Figure 4: Search forms of the IPI digital library (<https://www.literaturaspopulares.org/ipm/w/Consulta:Impresos>).

Several other libraries offer to explore their collections through this type of information as well, but in the form of indexes (Mapping Pliegos, Broadside Ballads Online, Comedias Sueltas USA, John Johnson Collection, Literatura Oral y Tradiciones Populares), as shown in **Figure 5**, or maps (Biblioteca Virtual Cordel), as exemplified in **Figure 6**. In the case of CBDRS, this editorial logic can be seen in the very organization of the library, where all the editions of the same title are grouped together. It is the chapbook as a multiple document that is emphasized here (Baena, Casas, and Cuadros 2013; Pena Sueiro and García 2014).

<h2>Índice impresores</h2>
Abadal, Ignacio
Administración de la Galería Literaria
Aguado, Afrodisio
Aguayo, Adolfo
Albarracín
Albert, A.

Figure 5: Index of printers (Mapping Pliegos, <https://web.archive.org/web/20220614122724/http://biblioteca.cchs.csic.es/MappingPliegos/indiceimpresores.php>).

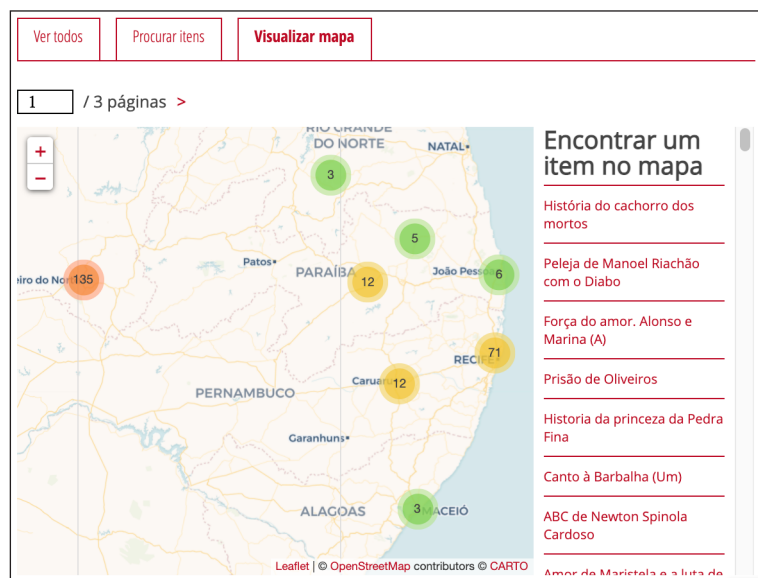


Figure 6: Map of printing places (Biblioteca Virtual Cordel, https://web.archive.org/web/20210508224202/http://cordel.edel.univ-poitiers.fr/geolocation/map/browse?sort_field=Dublin+Core%2CTitle).

The representation of chapbooks as documents is similar across the studied libraries. It seems to reflect the consistency of the production and material form of chapbooks, which then appear as an editorial format or genre. However, this tacit consensus breaks down when libraries focus on text and illustrations, two approaches specific to thematic libraries and portals.

3. Chapbooks and text: When the reader comes into play

Among the projects that focus on textual aspects, three solutions can be considered to put “users-readers” in contact with texts and not only with documents.

The first solution is to include in the description thematic keywords related to the type of text or the addressed themes. These metadata allow projects to generate indexes on the same model as those proposed for the names of printers or the publication places. However, the projects are faced with the difficulty of choosing a taxonomy to describe the themes or the nature of texts that are often specific to this type of printed object. Some, such as the John Johnson Collection or the Cambridge University Digital Library (CUDL), use a library-specific controlled vocabulary (Library of Congress Subject Headings) and “in-house taxonomies” to successfully capture all facets of these prints (Lambert 2015a).

The use of a pre-existing controlled vocabulary, primarily designed for the description of documentary texts, is challenging because the themes of the chapbooks are mainly related to fiction. This forces projects to multiply keywords to describe the

topic addressed by the texts, or to select only a few at the risk of overlooking some aspects (Nieto 2015). Other libraries, such as *Literatura de cordel y teatro* or *Mapping Pliegos*, then choose to create their own taxonomy, with the disadvantage, however, of a splintering of description methods from one project to another, making their interoperability difficult.

Libraries specialized in the study of one type of content can propose more detailed metadata specific to their corpus. For instance, *Complaintes Criminelles* gives some details on the crimes related by the *complaintes*, including links to newspapers that reported the facts. This emphasis on text and its content in the descriptions is taken to its extreme with *The Word on the Street* library, where each chapbook is described not with a list of metadata, but with a paragraph that looks like a scholarly commentary (see **Figure 7**). It contextualizes the text by providing information about the origin of the story being told, its form, as well as its reception and purposes. It insists more on the historical and literary aspects than on the documentary ones. By this choice of presentation of metadata, the project looks like a digital edition, where each chapbook benefits from a critical introduction and a transcription.

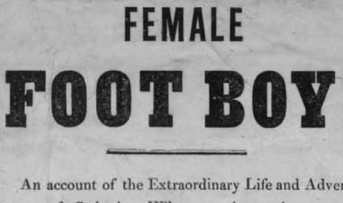
<p>Commentary</p> <p>This account begins: 'FEMALE FOOT BOY! / An account of the Extraordinary Life and Adventures of Catherine Wilson, an interesting young woman, about twenty years of age, daughter of respectable parents, near Perth, who assumed man's apparel at the age of fourteen, and hired herself to a drover, when she came to Edinburgh, and got into a respectable gentleman's family as a foot boy'. This broadside was printed by R. Reynolds, 489 Lawnmarket, Edinburgh.</p> <p>According to this account, Catherine Wilson assumed the name of John Thomson during her time disguised as a man. The creation of an alter ego was apparently in response to losing her parents at the age of six or seven. Whilst this story was most probably considered a curiosity and provided entertaining reading at the time, for today's audience it highlights the plight of women during the eighteenth and nineteenth centuries. In a society where women's civil, legal and political rights were almost non-existent, Catherine Wilson was forced to live as a man to earn her living and, as a lone woman, ensure her safety.</p> <p>Broadsides are single sheets of paper, printed on one side, to be read unfolded. They carried public information such as proclamations as well as ballads and news of the day. Cheaply available, they were sold on the streets by pedlars and chapmen. Broadsides offer a valuable insight into many aspects of the society they were published in, and the National Library of Scotland holds over 250,000 of them.</p> <p>View Transcription Download PDF Facsimile</p>	<p>Probable date published: 1820- shelfmark: L.C.1268</p> 
---	--

Figure 7: Scientific commentary of a broadside (*The Word on the Street*, <https://web.archive.org/web/20220614123703/https://digital.nls.uk/broadsides/view/?id=15929>).

These metadata about the text constitute a first level of analysis, which guides users in their research and gives them a general idea of the library collections. To get to the heart of the text, some projects provide users with full text search functionalities. These functionalities appear at two levels: first at the content level (*Fundación Joaquín Díaz*, *Biblioteca Virtual Cordel*, *McGill Library*) with a simple search bar associated with a visualization tool, then at the library level, that is, in all the collections (*Fundación Joaquín Díaz*, *IPI*, *McGill Library*).

While in some cases users can only view extracts of transcriptions that appear in the list of results of a query, in other cases the transcriptions are displayed directly to users in their entirety. This is the case for *CUDL* and *The Word on the Street*, which show the

transcription and the facsimiles of a content in front of each other. As for IPI, only a few prints have been transcribed and edited. It is a collaborative work in progress, which relies on the willingness of users to contribute. The results of this process constitute a collection of their own within the library. This strategy corresponds to the logic of fragmentation of the interface adopted by this project to adapt to different communities of users.

With these first two solutions, libraries provide users with services that incite them to do their own research and analysis of texts. They appear as tools, as workspaces that encourage the exploration of contents and adapt to the needs of users. However, digital libraries can choose a third way by themselves proposing an analysis of the content of chapbooks with a specific point of view, which implies preliminary work on the corpus.

The analysis of printed material by the library can first be part of its set of services, with virtual exhibitions or thematic maps, for example. The *Complaintes Criminelles* library displays a map on its home page that shows the places where the crimes described in the *complaintes* took place (see **Figure 8**). Similarly, the IPI library offers several maps, including a geolocation of the places where the events depicted in the prints took place, and a thematic classification with a colour code according to the type of these events. The digital library is thus a place for publishing research results, which in turn become tools that allow users to explore the corpus from a particular perspective.

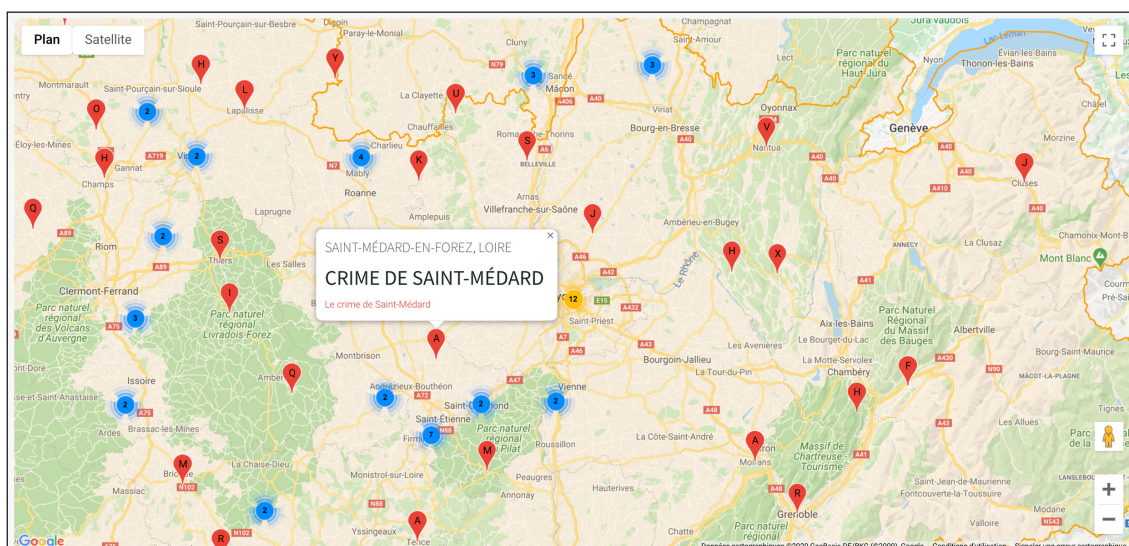


Figure 8: Map of the crimes depicted in the text (*Complaintes Criminelles*, <https://web.archive.org/web/20220116164555/https://complaintes.criminocorpus.org/>).

In the case of exhibitions or maps, only a part of the library is dedicated to highlight the results of a scientific research. However, it is possible to consider a library organized solely from this point of view, as shown by the project *Literatura de cordel y teatro en España*. The latter is the direct result of a doctoral thesis whose objective was to study chapbooks derived from plays. It makes the textual approach the core of its project. This leads to an organization that is very different from the other libraries in our study, insofar as the structuring unit of the library is no longer the print, but the play. A single work can thus group together several chapbooks from different printers' workshops and different time periods.

The *Literatura de cordel y teatro* project is an extreme case in the choice of a textual approach to chapbooks, made possible by the restriction of the corpus to a specific genre. This emphasis on the literary aspect of the prints can be seen in the very title of the project. Indeed, while most of the projects emphasize the document, its form, or its production (Mapping *Pliegos*, Biblioteca Virtual *Cordel*, *Impresos Populares Iberoamericanos*, McGill's *Chapbook Collection*, *Comedias Sueltas USA*, *Complaintes Criminelles*, *Broadside Ballads Online*), it is the textual aspect that is emphasized here (*Literatura de cordel y teatro*), something that can also be found with the project *The Word on the Street*.

Thus, though portals and general libraries only offer metadata related to the general theme or type of text, services and detailed metadata on the content of printed materials are more common in specialized libraries. The same can be said for an approach of chapbooks from the point of view of their illustrations.

4. Chapbooks and illustrations: Some digital quivering

The study of illustrations, which are part of the visual identity of chapbooks, is found at the margin of digital libraries. It appears in the metadata, usually with an indication of the number of illustrations (Mapping *Pliegos*, CUDL, Biblioteca Virtual *Cordel*, McGill Library) or a brief description (Fundación Joaquín Díaz). However only four of the seventeen projects studied have specific services dedicated to illustrations in addition to these few mentions in the metadata (Mapping *Pliegos*, John Johnson Collection, *Broadside Ballads Online*, and *Comedias Sueltas USA*).

These four projects have taken a different approach to describing and organizing illustrations. On the Mapping *Pliegos* side, illustrations, of which only a selection are accessible, are grouped according to their similarities based on a distinguishing feature, such as "women with a fan" or "men playing a musical instrument" (see **Figure 6**). The

John Johnson Collection and Broadside Ballads Online libraries have chosen a thematic classification based on a specialized thesaurus for describing images called IconClass. Finally, Comedias Seltas USA organizes its collections of illustrations by stamp; occurrences of the same stamp are grouped together.

These approaches are reflected in the very form of the catalogues and in the choice of the adopted techniques. To highlight visual similarities between illustrations, Mapping Pliegos uses the social network Pinterest and its wall of imageboards, which makes it possible to identify recurring motifs at glance (see **Figure 9**). Although John Johnson Collection and Broadside Ballads Online also take a thematic perspective, it is presented as an index of concepts derived from IconClass. Finally, Comedias Seltas USA has considered a catalogue with a list of prints where the stamp has been used, which is an ongoing project.



Figure 9: Some illustrations of Mapping Pliegos on Pinterest (<https://web.archive.org/web/20210128121854/https://www.pinterest.es/bibtntcsic/grabados-mapping-pliegos/>).

These technical solutions target different research objectives. Mapping Pliegos, the John Johnson Collection, and Broadside Ballads Online provide tools for exploring chapbooks through their illustrations. These tools make it possible to analyze some themes or thematic relationships that exist between a text and an illustration. For the Comedias Seltas library, the categorization of illustrations by stamp is a method of dating the chapbooks or identifying printers. It also allows users to study the way an illustration is distributed in a corpus, and how it was used. In this perspective, the Broadside Ballads Online library offers, in addition to its thematic index, a search engine by images (see **Figure 10**). Users can select all or part of an illustration and obtain identical or similar results, thus sharing the objectives of Comedias Seltas in terms of analysis of the use and diffusion of stamps and patterns in a corpus.

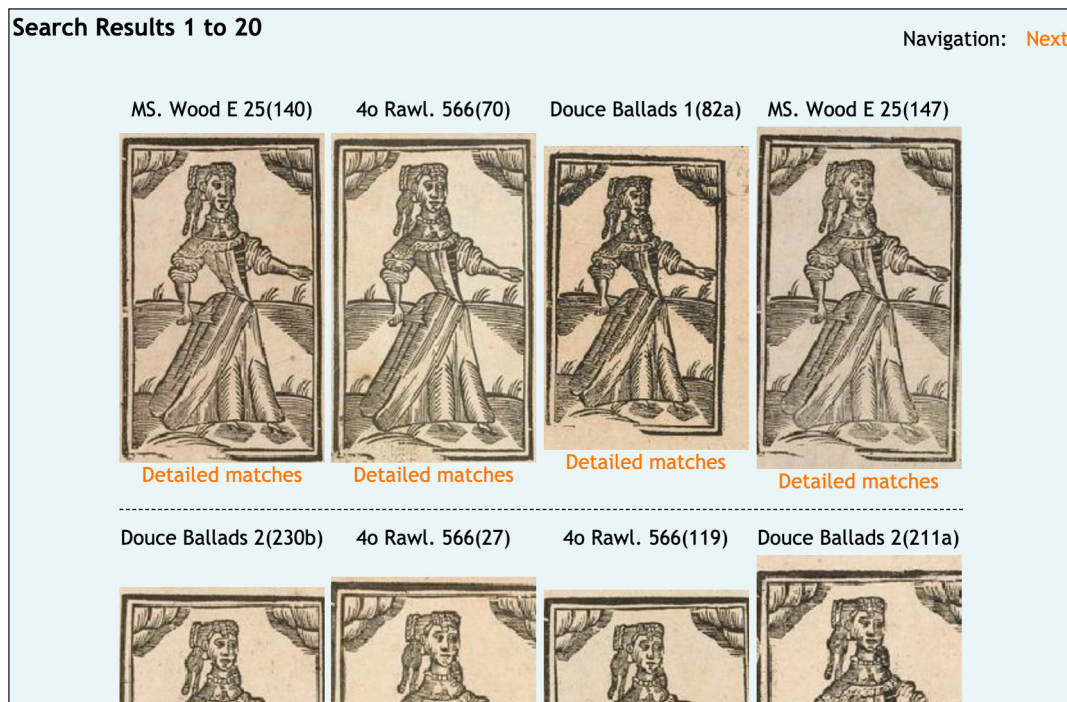


Figure 10: Results of a query by image (Broadside Ballads Online, <https://web.archive.org/web/20201204091345/http://imagematch.bodleian.ox.ac.uk:8000/>).

Conclusion: Digital libraries facing the challenge of modelling a pluridisciplinary object

Chapbooks are complex documents that, despite their apparently simple form, have many facets. These facets lead to multiple interpretations, whether from the point of view of book history, literature, or art history. This pluridisciplinarity appears to be one of the main challenges that the projects must face. If some can avoid this issue by focusing on a specific aspect (*Literatura de cordel y teatro* with a literary approach, *Complaintes Criminelles* with a historical and musicological approach, *The Word on the Street* with a historical approach), others multiply search systems to address the needs of a diverse public (IPI, Mapping Pliegos, Broadside Ballads Online, *Comedias Seltas USA*).

Despite an impression of abundance in the selected metadata and services, the projects tend to focus on the documentary and material aspects of chapbooks through an important work on bibliographical metadata. This is a common thread that gives structure and consistency to the representation of chapbooks in digital libraries. The study of chapbooks from the viewpoint of their editorial process constitutes indeed a preferred approach in the field of study on this type of content (Gomis and Botrel 2019). This is reflected by the way the digital libraries are structured.

Few of these libraries have yet been dedicated to an in-depth study of texts and illustrations, although some initiatives in this direction should be noted. The content of a print, which is another important line of research in the traditional analysis of chapbook literature (Gomis and Botrel 2019), is most often revealed through keywords describing the type of text and general topic. The projects here face the difficulty of choosing a pre-existing taxonomy that would suit not only books, but also non-fiction.

When the transcription is available, it is mainly used for full-text search, and is most often “invisible” to the user, who is only presented with portions of transcribed text during a search. Only *Literatura de cordel y teatro* and *The Word on the Street* offer an analysis of texts, either through a scientific commentary or an organization of the *pliegos* by the work of which they are a rewriting.

This absence of text and illustration in most projects can be explained by the mass of data they are confronted with. This amount of data makes it difficult to carry out further analysis without requiring advanced digital techniques that are costly from human, temporal, and financial points of view. However, recent advances in optical character recognition or computer vision applied to heritage documents open new perspectives. We can think of the processing and exploration of large corpora of chapbooks, whether in terms of text mining, with the possibility of identifying rewrites or studying the circulation of information on a large scale, or even of analysis of illustrations and of their reuse, as demonstrated by the *Broadside Ballads Online* project and more recently by the National Library of Scotland with its experiments on similarity detection within large corpora of chapbooks’ engravings (Dutta, Bergel, and Zisserman 2021).

Index of cited projects

- Biblioteca Virtual Cordel: <http://cordel.edel.univ-poitiers.fr/>
- Broadside Ballads Online: <http://ballads.bodleian.ox.ac.uk/>
- Catálogo y Biblioteca Digital de Relaciones de Sucesos (CBDRS): <https://www.bidiso.es/CBDRS/>
- Calaix: <http://calaix.gencat.cat/handle/10687/120608>
- Colección Literatura oral y Tradiciones populares: <http://www.bibliotecanacionaldigital.gob.cl/bnd/627/w3-channel.html>
- Comedias Sueltas USA: <https://comediassueltasusa.org/>
- Complaintes Criminelles: <https://complaintes.criminocorpus.org/>
- Cambridge University Digital Library (CUDL) – Spanish Chapbooks: <https://cudl.lib.cam.ac.uk/collections/spanishchapbooks/1>
- Digital Public Library of America: <https://dp.la/>
- Europeana: <https://www.europeana.eu/fr>
- Fundación Joaquín Díaz: <https://funjdiaz.net/pliegos-listado.php>
- Impresos Populares Iberoamericanos: <https://www.literaturaspopulares.org/ipm/w/Inicio>
- IconClass: <http://www.iconclass.org/help/outline>
- John Johnson Collection: <https://www2.bodleian.ox.ac.uk/johnson>
- Library of Congress Subject Headings: <https://id.loc.gov/authorities/subjects.html>
- Lira Popular: https://bibliotecadigital.uchile.cl/discovery/search?vid=56UDC_INST:56UDC_ABELLO
- Literatura de cordel y teatro: <http://www.pliegos.culturaspopulares.org/>
- Mapping Pliegos: <http://biblioteca.cchs.csic.es/MappingPliegos/>
- McGill Library's Chapbook Collection: <https://digital.library.mcgill.ca/chapbooks/index.php>
- RUBI: <http://rubi.casaruiarbosa.gov.br/handle/fcrb/517>
- The Word on the Street: <https://digital.nls.uk/broadsides/index.html>

Competing interests

The author has no competing interests to declare.

Contributions

Editorial

Section Editor

AKM Iftekhar Khalid, The Journal Incubator, University of Lethbridge, Canada

Gimena del Rio Riande, The Journal Incubator, Universidad de Buenos Aires, Argentina

Copy Editors

Christa Avram, The Journal Incubator, University of Lethbridge, Canada

Morgan Pearce, The Journal Incubator, University of Lethbridge, Canada

Layout/Managing Editor

Virgil Grandfield, The Journal Incubator, University of Lethbridge, Canada

Bibliography

Baena, Francisco, Inmaculada Casas, and Maria Teresa Cuadros. 2013. "Una biblioteca digital de prensa antigua. El caso de las relaciones de sucesos sevillanas (siglo XVI-XVIII)." *Revista Internacional de Historia de La Comunicación* 1(1): 1–32. Accessed October 27, 2022. <http://hdl.handle.net/11441/24774>.

Bermès, Emmanuelle. 2011. "La gestion des collections numériques: Vers une numérisation durable." In *Manuel de numérisation*, edited by Thierry Claerr and Isabelle Westeel, 287–301. Paris: Éditions du Cercle de la Librairie.

Botrel, Jean-François. 2000. "Littérature et imprimés de cordel dans la péninsule ibérique." In *Des conquêtes de Charlemagne au Brésil. Le Moyen Âge européen dans la littérature populaire brésilienne. Catalogue de l'exposition*, edited by R. Lemaire and A. Moreau. Poitiers: Médiathèque François Mitterand. Accessed August 08, 2022. https://botrel-jean-francois.com/Cordel_colportage/Lit._cordel.html.

———. 2001. "El Género de Cordel." In *Palabras para el pueblo. I. Aproximación general a la literatura de cordel*, edited by Luis Díaz G. Viana, 41–69. Madrid: CSIC. Accessed August 08, 2022. http://www.cervantesvirtual.com/obra-visor/el-gnero-de-cordel-0/html/0133d94a-82b2-11df-acc7-002185ce6064_7.html#I_0_.

Dutta, Abhishek, Giles Bergel, and Andrew Zisserman. 2021. "Visual Analysis of Chapbooks Printed in Scotland." In *The 6th International Workshop on Historical Document Imaging and Processing (HIP '21), September 5–6, 2021, Lausanne, Switzerland*. New York: ACM. Accessed August 08, 2022. <https://www.robots.ox.ac.uk/~vgg/research/chapbooks/data/dutta2021visual.pdf>. DOI: <https://doi.org/10.1145/3476887.3476893>.

Gomis, Juan, and Jean-François Botrel. 2019. "'Literatura de Cordel' from a Transnational Perspective. New Horizons for an Old Field of Study." In *Crossing Borders, Crossing Cultures*, edited by Massimo Rospocher, Jeroen Salman, and Hannu Salmi, 127–142. Berlin: De Gruyter Oldenbourg. DOI: <https://doi.org/10.1515/9783110643541-008>.

La Polola, o, La reina de mis cantos: canciones: arpa, guitarra i piano. 1908. In: Becerra, F. M. (Ed.) Santiago de Chile: Imprenta i Encuadernación Universitaria. Accessed October 27, 2022. <http://www.bibliotecanacionaldigital.gob.cl/bnd/627/w3-article-616974.html>.

Lambert, Julie Anne. 2015a. "L'embarras du choix? Complexité & perplexité dans le catalogage des éphémères de la John Johnson collection." *Fabula Colloques. Les éphémères, un patrimoine à construire*. Accessed August 08, 2022. <https://www.fabula.org:443/colloques/document2890.php>.

———. 2015b. "Pérenniser l'éphémère: paradoxe et défi dans le contexte de la John Johnson Collection." *Fabula Colloques. Les éphémères, un patrimoine à construire*. Accessed August 08, 2022. <https://www.fabula.org:443/colloques/document2940.php>.

Leblanc, Elina. 2019. "Bibliothèques numériques enrichies et participatives: Utilisateurs, Services, Interfaces." PhD diss., Université Grenoble Alpes (ComUE). Accessed August 02, 2022. <https://tel.archives-ouvertes.fr/tel-03784540v1>.

Mion-Mouton, Fanny. 2012. "Bibliothèques numériques et coopération: comparaisons internationales." *Mémoire d'étude DCB 21*. Villeurbanne: Enssib.

Nieto, Philippe. 2015. "Cataloguer les éphémères. Quelques pistes de réflexion." *Fabula Colloques. Les éphémères, un patrimoine à construire*. Accessed August 02, 2022. <https://www.fabula.org:443/colloques/document2896.php>.

Pena Suerio, Nieves, and Sandra García Álvarez. 2014. "El catálogo y biblioteca digital de relaciones de sucesos: bases de datos bibliográficas, textos e imágenes." In *Humanidades Digitales: desafíos, logros y perspectivas de futuro*, edited by Sagrario López Poza and Nieves Pena Sueiro, 335–345. A Coruña: Universidade da Coruña.

Pierazzo, Elena. 2015. *Digital Scholarly Editing: Theories, Models and Methods*. Farnham: Ashgate. Accessed August 02, 2022. <https://hal.archives-ouvertes.fr/hal-01182162/document>.

Rollin, Charles M. 1795. *Historia de los emperadores romanos desde Augusto hasta Constantino*. Madrid: En la Oficina de Don Blas Roman. Accessed December 07, 2022. <http://www.bibliotecanacionaldigital.gob.cl/bnd/644/w3-article-618554.html>.

Trobos discretos y divertidos, para entretener a los curiosos y cantar los galanes á sus queridas damas, y especialmente los que se sienten agraviados por no ser correspondidos. 1858. Edited by José María Moreno. Accessed October 27, 2022. <https://zenodo.org/record/6483452>.

Vitali-Rosati, Marcello. 2016. "What Is Editorialization?" *Sens Public*, January. Accessed August 01, 2022. <http://sens-public.org/article1059.html>. DOI: <https://doi.org/10.7202/1043408ar>

