

Tomasz Neugebauer¹, Corina MacDonald², Felicity Tayler³

Artexte Metadata Conversion to EPrints: Adaptation of Digital Repository Software to Visual and Media Arts Documentation

1. *Digital Projects and Systems Development Librarian at Concordia University*
2. *Project Manager at Artexte*
3. *Information Specialist at Artexte*

The final publication is available at www.springerlink.com (DOI: [10.1007/s00799-011-0077-5](https://doi.org/10.1007/s00799-011-0077-5))

Abstract:

Purpose: The following research questions structured our analysis: Does an open access institutional repository model respond to the needs of a non-academic documentation centre? Is EPrints software a good match to support the needs of the existing metadata describing Artexte's collection? What are the customizations required to accommodate existing Artexte metadata using EPrints?

Methods: We exported the existing metadata schema and sample data in Artexte's three databases, performed a manual evaluation of metadata quality and compared the 49 Artexte fields to those available within the EPrints schema.

Results: We identify the metadata elements that mapped by default without the need for customization or modification and those which would need to be added to EPrints using configuration files. We also identify the custom software development to accommodate Artexte metadata using EPrints: the bilingual controlled vocabulary demands an extension of the EPrints subject taxonomy model with thesaurus semantic relationships.

Conclusions: Comparing Artexte and EPrints metadata schemas, we found that 15 out of 49 fields mapped by default without need for modification, 25 fields would need to be added to EPrints configuration files and 1 field will be removed during the migration. With only 8 fields requiring some special attention, we conclude that EPrints is suitable to the needs of Artexte's bibliographic data management.

Keywords: *Metadata, EPrints, Art Documentation, Open Access*

Artexxe Metadata Conversion to EPrints: Adaptation of Digital Repository Software to Visual and Media Arts Documentation.

1. Introduction

Artexxe is an independent organization founded in 1980 with a mandate to support research, interpretation, and dissemination activities contributing to the appreciation of contemporary visual arts in Canada. The primary users of the organization's services are a national and international community of artists, art historians and curators although it also serves art enthusiasts in the general public. The centre has extensive holdings of paper-based documents such as exhibition catalogues, periodicals and ephemera in artists' files. These materials represent a historic body of literature within a discipline that has remained tied to a paper paradigm for publishing secondary sources in the field. Growing interest in digital publishing platforms prompted Artexxe to look to an open source solution that would allow existing bibliographic data, describing the physical collection, to be complimented by a future collection of digital objects. As our user groups tend to overlap with the publishers in the field, the self-archiving function of EPrints [33] as well as its firm position within the area of open access institutional repositories, confirmed our interest in evaluating its suitability to Artexxe's collection and bibliographic data management needs.

The following research questions structured our analysis: Does an open access institutional repository model respond to the needs of a non-academic documentation centre? Is EPrints a good match to support the needs of the existing metadata describing Artexxe's collection? And finally, what are the customizations required to accommodate existing Artexxe metadata using EPrints?

This paper describes and evaluates the current metadata of the Artexxe databases and provides a detailed comparative analysis with the metadata schema of EPrints. We discuss the applicability of open access to Artexxe, and the rationale for selecting the EPrints software, over other options, for this detailed analysis. Our findings focus on the description of customization, configuration and development that would be necessary to migrate Artexxe's existing metadata to EPrints.

Throughout the discussion, we will refer to examples of materials in the collection that document the practice of Gabor Szilasi. Szilasi is a photographer born in Budapest, Hungary, in 1928, immigrating to Canada in 1957, and finally settling in Montreal in 1959. Before his arrival in Canada, Szilasi documented images of the Hungarian Revolution, but his photography is most widely recognized for its depiction of rural Québec regions and the vernacular architecture of Montréal.

2. Literature Review

This study is unique in that it evaluates the suitability of a digital repository software's metadata model based on the analysis of an existing metadata catalogue for collections of documents in the Fine Arts. Related literature includes the evaluation of metadata quality [3,15,17,24], the evaluation of digital repository software [6,7,10,21,23,27,29,31,32,36], the migration of metadata [1,11,17,35], metadata and digital repository software in the Fine Arts [14,30], and open access [4,5,8,9,12,16,20,25,26]. As the analysis in this paper is based on the existing collection at Artexxe and the system it has developed to support its users' needs, there is also a brief discussion of the literature around the history and services of the organization [2,18,19].

Sets of descriptive and structural metadata specifications are fundamental concepts in the study of digital library structure [13]. The metadata catalogue and the services of indexing, searching

and browsing, are based upon a digital library's metadata. A descriptive analysis of metadata that is tailored for a particular user community is a valuable contribution as a case study of digital library structure in the Fine Arts. As IFLA reminds us, metadata enables the user to find, identify, select and obtain resources specific to their disciplinary needs [15]. Furthermore, an analysis resulting in a description of the configuration and customization that would be necessary to migrate this metadata to an open access digital repository software used for open access self-archiving in academia reveals insights into the versatility and interoperability of the open access digital repository software.

Two recent reports evaluating digital repository software are related to our study in that they consider EPrints as candidate software. The National Library of Medicine evaluated ten systems, both open source and commercial, and ultimately selected Fedora for further pilot testing and listed among the reasons its "excellent underlying data model that can handle NLM's diverse materials" [21]. Similarly, London School of Economics recently announced their intention to "implement of Fedora at the core" of their digital library, listing among the reasons Fedora's superior "data model to support complex object types" [10]. However, neither of these reports provides the type of detailed analysis of metadata that we present in this paper since they do not provide details to demonstrate which particular metadata properties of the "data model" cause a challenge in accommodating the metadata of an existing collection.

There is also a growing literature describing workflows for batch importing into digital repository software [35], including tutorials for batch ingestion of ProQuest ETD metadata into Digital Commons platform using XSLT [1] and the procedures for bulk importing into DSpace [11]. However, these reports are focused on the metadata domain of electronic theses and dissertations and academic journal articles. While Artexte metadata includes magazine/journal articles and conference papers, it also adds to that: essays, exhibition catalogues, exhibition ephemera, artists' books, and audiovisual documentation that require customized metadata fields. Some of the "sociotechnical" (defined as the mutual constitution of social and technical processes) challenges with crosswalking and migration of metadata reported by Khoo & Hall [17] were also observed in our evaluation of metadata. Khoo & Hall list unique or specialized metadata fields that were a result of sociotechnical considerations such as: former title, sort title, acronym, alternate title and alternate spelling. Khoo & Hall [17] also identify the lack of strict application in practice of the use of controlled subject vocabulary, causing inconsistency in the completeness of metadata records over time.

The evaluation of metadata quality in digital libraries is equally a related research area in our consideration of metadata migration. Averkamp & Lee point out limitations of the bepress schema used by the Digital Commons platform, such as the incomplete support for Unicode and the lack of extensibility to include the discipline [1]. They also list other metadata quality limitations identified in their migration: the granularity of the publication date, and the institutional convention to use all capital letters in the title field [1]. Ochoa & Duval provide a review of quality evaluation studies for metadata, the majority of which take the approach to manually review a statistically significant sample of metadata against a set of parameters [24]. The Bruce & Hillman [3] definition of metadata quality measures (completeness, provenance, accuracy, conformance to expectations, logical consistency and coherence, timeliness, accessibility) can be used by aggregators as "hints about where they might look for trouble in legacy and multiple-source data". In this study, we used the Bruce & Hillman quality measures but we focused only on identifying those metadata quality issues that would cause problems in the migration to EPrints.

In the study most closely related to our own disciplinary area, Gray describes the Kultur project, a JISC funded research project to explore self-archiving in creative and applied arts, as a solution for giving access to academic research in this field [14]. In addition to proposing “art/design items” as a document type, Gray reports on two additional item types that were added to the metadata as “mediums of research dissemination”: “exhibitions” and “performance” [14]. Visual aspects of user interface design are also important, as the practitioners in this field are concerned how interface design “would reflect upon their actual work” [14]. In addition, Gray also identifies other areas that should be the subject of future investigation, such as copyright management in a field where research output has both an academic and commercial function [14]. Although related to work presented here, there are significant differences between the two methodologies. Gray’s study for the Kultur project was based on a survey of researchers, one-on-one interviews, and the development of a demonstrator repository through usability testing [14], whereas our study describes the suitability of a digital repository based on the analysis of an existing metadata collection.

Beyond differences in methodologies, there is a critical distinction between the types of material collected in the Kultur project and those of interest to Artexte. Kultur attempts to capture the primary material (i.e., the art works themselves), as well as the secondary sources that document or discuss it critically in terms of reception. The example given is moving images which may function “simply as documentation of a specific event such as a conference or a performance, or it may be the actual research practice itself, such as a video work” [14]. In the case of Artexte, the objective would be the collection of secondary sources as the collection development policy has traditionally precluded the acquisition of works of art. In the project’s final report on metadata, Sheppard summarises these development efforts for the metadata schema and deposit workflows created as part of the project [30].

As a reflection upon research patterns specific to creative disciplines, Gray confirms that “documentation of a particular piece of arts research could also involve digital images, audio or video, posters, articles or text” [14]. The recognition of this diversity is a principle that Artexte has applied to its collection development since inception. In her history of the organization, Léger describes the mandate of Artexte to collect, organize and disseminate information reflecting all aspects of contemporary visual arts since 1965 with a particular emphasis on Québec and Canada [19]. Léger elaborates upon the diversity of the material that makes up the information sources and the corresponding flexibility of Artexte’s cataloguing practices [19]. In a similar sense, Gray concludes that “the actual way that art research is valued is often outside the traditional academic structure and these sites of critical exploration can often be transitory” [14]. This corresponds to the status of Artexte as an independent organization with a subject-specific collection of material, Bradley and Johnstone explain that when the centre was founded in 1980, it acted as a point of centralization for Canadian exhibition catalogues and ephemera, art magazines and other forms of publishing that reflected a growing critical discourse from multiple sectors including museums, galleries and artist-run centres “for which neither a resource centre nor a means of distribution and cataloguing had previously existed” [2].

The metadata that structures Artexte’s bibliographic databases today is a reflection of the cataloguing rules developed to create the *Catalogue of Catalogues (1980-2004)*, a bibliographic tool produced biannually that listed the publications collected and distributed by Artexte. The first issue included 505 titles produced between 1968 and 1981 by 29 museums across the country. Until the distribution service ended in the mid-nineties, the scope and breadth of international publishing in the arts represented in the *Catalogue*, grew exponentially. This cataloguing practice was eventually automated in the mid 1980s providing the basis for the current bibliographic database. Latour and Tayler discuss more recent changes in research and publishing trends in the field, reflecting upon an increased expectation from researchers to find content online and a corresponding need for greater web presence for research and publishing in contemporary Canadian art [18]. As a solution, they suggest the exploration of open-source, interoperable digital collections or bibliographic database software platforms. Latour and Tayler observe that Artexte’s strength lies in its ability to reflect its user community by bringing together information produced by various sources, including research institutions (such as

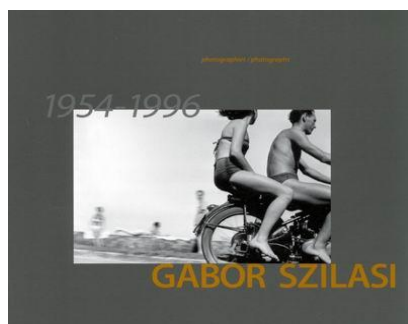
museums and public or university galleries), academic institutions, and independents (such as artists, private galleries, magazine publishers, and artist-run centres – as well as independent curators, critics and other cultural workers) [18], this ability to draw on a range of sources supports Gray's conclusion that "in a community that is as diverse as the arts, with its varying disciplines, cultures and outputs, a real advantage was to bring the evidence of this research together" [14].

3. Current Artexxe Information Organization and Retrieval

Artexxe's bibliographic data is currently held in three separate databases. The existing system is built using DB Text Works and loosely based on ISBD standards. These standards have been customized to accommodate a greater degree of detail in bibliographic description in response to user queries particular to the field of Fine Arts. The first bibliographic database contains records describing monographic publications in the collection (e.g. exhibition catalogues, anthologies, and conference proceedings) and audio visual formats (e.g. audio cassettes, audio CDs, DVDs and vinyl records). It also holds records for special document types such as artists' books (books made by artists as works of art). A second database holds records for periodicals, including the titles and holdings represented in the collection.

A third database, the files database, serves a dual purpose. Firstly, its records describe the classification system of the collection (unique to Artexxe). Secondly, it is linked to fields in the first two bibliographic databases in order to provide call numbers for items in the physical collection (COTD). For example, in the files database there exists a record to describe the classification 410 – SZILASI, GABOR. The numeric code 410 associates this file to the group "Canadian Artists" followed by the name of the artist described. A solo exhibition catalogue, such as *Gabor Szilasi : Photographs, 1954-1996* (Montréal ; McGill-Queen's University Press, 1997) would then be assigned 410 – SZILASI, GABOR as a call number in the bibliographic database thereby facilitating retrieval from the collection.

Figure 1. Record View for Gabor Szilasi Item in Artexxe Collection



DOC NO. 13265
 TITRE / TITLE **Gabor Szilasi : Photographies, 1954-1996 = Gabor Szilasi : Photographs, 1954-1996**

AUTEUR(S) / AUTHOR(S) Harris, David (auteur et rédacteur) // Michel, Franck (conservateur et auteur) Blouin, Marcel (préfacier) Langford, Martha (préfacier)

PUBLISHER(S) / ÉDITEUR(S) Montréal, Qc, McGill-Queen's University Press Vox Populi (1997).

DOSSIER / FILE **410 - SZILASI, GABOR**

COLLATION 144 p. : 96 ill. (17 en coul.) ; 21 X 23 cm TYPE catalogue d'exposition Exhibition catalogues monographie

COMMISSAIRE / CURATOR Michel, Franck

ÉVÈNEMENT / EVENT Exposition organisée par Vox Populi dans le cadre du Mois de la photo à Montréal édition 1997. - Dates d'exposition: 28 août - 2 nov. 1997, Musée des beaux-arts de Montréal.

LANG. FRE ENG
 ISBN / ISSN ISBN 0773517286
 GEOG. CODE EUR-HUN-BUDAPEST CAN-QC-MONTRÉAL

RÉSUMÉ Rétrospective de la photographie documentaire (ou d'auteur) pratiquée par l'artiste d'origine hongroise depuis son arrivée au Canada. Ce document reproduit une abondance d'oeuvres, regroupées selon 14 thèmes (la plupart géographiques), touchant le milieu rural autant qu'urbain, la place de l'architecture dans la vie collective et individuelle, et les ambiguïtés du genre portrait. F. Michel présente Szilasi comme une personnalité au regard humaniste, alors que Harris propose un itinéraire de l'oeuvre par étapes en faisant ressortir les valeurs véhiculées par les images. Dossier bibliographique 12 p.

ABSTRACT A retrospective of photographic works (documentary and artistic) produced by this Hungarian-born artist since his arrival in Canada. The document includes numerous reproductions gathered around 14 themes - primarily geographic - relating to urban and rural environments, the place of architecture in collective and individual lives, and ambiguity in the portrait genre. F. Michel introduces Szilasi as a personality with a humanist vision, while Harris surveys stages of the artist's work, foregrounding the values conveyed by the images. Bio-bibliography 12 p.

SUJET(S) / SUBJECT(S) [DOCUMENTARY PHOTOGRAPHY / PHOTOGRAPHIE DOCUMENTAIRE](#) // [PORTRAIT / ARCHITECTURE VERNACULAIRE / FAÇADES / HABITATIONS / INTÉRIEURS / PHOTOGRAPHIE DE RUE / PORTRAIT PSYCHOLOGIQUE / RURALITÉ QUÉBÉCOISE / VERNACULAR ARCHITECTURE / HOUSE\(S\) / INTERIORS / STREET PHOTOGRAPHY / PSYCHOLOGICAL PORTRAIT / QUEBEC RURALITY](#)

ARTISTE(S) / ARTIST(S) [SZILASI, GABOR](#)
 AUTRE / OTHER // [BUDAPEST / MONTRÉAL](#)

3.1 Controlled Vocabulary

Both the bibliographic and periodical databases contain fields with controlled vocabulary. For example, the contents of **subject descriptor** fields are chosen from a list of controlled vocabulary. The subjects are a predetermined list of terms specific to contemporary art practices. The field can contain multiple entries. Artexte provides subject analysis in French and English if the publication itself is bilingual. In our example, *Gabor Szilasi : Photographs, 1954-1996*, the subject descriptors attributed to the exhibition catalogue are listed in Table 1.

Table 1. Subject Descriptors for Gabor Szilasi Item in Artexte Collection in English and French

English	French
DOCUMENTARY PHOTOGRAPHY	PHOTOGRAPHIE DOCUMENTAIRE
VERNACULAR ARCHITECTURE	ARCHITECTURE VERNACULAIRE FAÇADES
HOUSE(S)	HABITATIONS
INTERIORS	INTÉRIEURS
STREET PHOTOGRAPHY	PHOTOGRAPHIE DE RUE
PSYCHOLOGICAL PORTRAIT	PORTRAIT PSYCHOLOGIQUE
QUEBEC RURALITY	RURALITÉ QUÉBÉCOISE

PORTRAIT	BUDAPEST / MONTRÉAL
----------	---------------------

A separate subject field identifies the **artist as subject**; the artists' names in this field are also part of a controlled vocabulary list of personal names. In our example, this field holds the name SZILASI, GABOR. If it was the catalogue of a group exhibition (including more than one artist) then the names of the other artists would be represented by multiple entries.

3.2 Multiple Entry Fields

In addition to artist names as subject, other fields can hold multiple entries. The publisher field, for example takes into consideration the trend towards co-publishing between museums and galleries. In our sample record, there are two publishers, both situated in Montreal:

- McGill-Queen's University Press
- Vox Populi

Also of note is the **document type** field which holds multiple entries as documents often need to be described as having more than one function, or a form and a function. For example, ephemera such as a poster that announces an exhibition may also contain additional information of interest to our researchers (such as an introductory text). As a result, the ephemera shown in Figure 2 would be entered into the database with the same level of description that a substantial exhibition catalogue would receive. In this case, the document type field would contain two entries: EXHIBITION EPHEMERA (to identify it as a document issued alongside an exhibition) and FLYERS, POSTERS to identify its physical form.

Figure 2. Item Catalogued with Multiple Document Types in Artexte Collection

DOC NO. S039

TITRE / TITLE **City's Heritage : A Photographic Memory : An Exhibiton Organized by L'Association photographique de Montréal**

AUTEUR(S) / AUTHOR(S) Arbour, Madeleine (author) Bumbaru, Dinu (author) Hébert, Robert (artist and author)

PUBLISHER(S) / MONTRÉAL, Qc, Conseil des arts de la Communauté urbaine de Montréal ([1989]).

DOSSIER(S) / FILE(S) **313 - COMMUNAUTÉ URBAINE DE MONTRÉAL - CONSEIL DES ARTS**

COLLATION 1 poster : 11 ill. ; 28 x 22 cm (unfolded) TYPE Flyers, posters Exhibition ephemera Exhibition catalogues poster : 87 x 28 cm

ÉVÈNEMENT / EVENT Part of the "Exposer dans l'île" programme. LANG. ENG
- Exhibition dates: 12 - 29 Oct. 1989, GEOG. CODE CAN-QC-MONTREAL
Maison de la culture Marie-Uguay
(Montréal) ; 17 - 26 Nov. 1989, City Hall
(Ville Saint-Pierre) ; 15 Dec. - 21 Jan. 1990,
Centre culturel de Pierrefonds ; 31 May - 20
June 1990, Centre culturel de Dorval.

ABSTRACT Includes brief introductory texts addressing the links between the 150th anniversary of the invention of photography, the urban architectural landscape of Montréal and the importance of a photographic record of the city. Biographical notes on the ten participating artists.

SUJET(S) / SUBJECT(S) // INVENTION OF PHOTOGRAPHY

ARTISTE(S) / ARTIST(S) MILLER, DAVID FOURNIER, ROBERT RUWEDEL, MARK GUTSCHE, CLARA RAJOTTE, NORMAND SZILASI, GABOR LAFOREST, ALAIN BOULET, MICHEL HEBERT, ROBERT MERRETT, BRIAN

AUTRE / OTHER // DE LOTBINIERE, Seigneur GASPARD JOLY

NOTES Also published in French under the title: "Une ville en héritage : Mémoire photographique : Une exposition de l'Association photographique de Montréal".

3.3 Metadata quality

In preparation for the work of mapping Artexte's current fields to the EPrints schema, a manual evaluation of select Artexte metadata elements was conducted to gauge the quality of the current metadata. Using Bruce and Hillman's seven characteristics of quality metadata as a guideline, this process helped in the identification of problems that would impact the mapping of Artexte metadata to EPrints and the resulting migration process [3]. Table 2 describes our findings with regards to each of these seven characteristics.

Table 2. Evaluation of Artexte metadata quality based on Bruce and Hillman's 7 characteristics

Provenance	Cataloguing at Artexte has always been done by a small number of librarians, or by interns and art historians trained and supervised by librarians. Thus the provenance of metadata in the catalogue has remained reliable over the years.
Accuracy	Artexte librarians follow well established cataloguing rules and practices which minimize accuracy problems. A cataloguing guide was developed to provide detailed instruction and examples for each field, minimizing inconsistencies.
Completeness	Artexte metadata has been adapted over the years to provide an appropriate level of description for users of the collection. Over time elements have been added and removed from the set to reflect the changing nature of the collection and user needs.
Logical consistency and coherence	<p>The logical consistency of Artexte metadata is good overall; the high level of conformance to the rules of the cataloguing guide has been effective in maintaining consistency over time.</p> <p>There is one field however, Geographic code, that has shifted in meaning over time and is no longer coherent as an access point for searching or browsing. According to the cataloguing guide this field relies on a controlled vocabulary to designate the birthplace of an artist. However, because this field is used to describe objects in the collection, it has shifted in meaning over the years to be used as a geographic subject field describing the object rather than the artist. The inconsistent use of the field results in metadata poor quality and it will be removed in the mapping to EPrints.</p>
Timeliness	Almost one quarter of the records in the bibliographic database (approximately 4500 of 22000 records) are currently classified as 'in treatment', meaning they have not yet been assigned call numbers and have only accession level metadata. This is the result of a cataloguing time lag that has accrued due to limited resources available for cataloguing. Despite this lag it was deemed important to make these records accessible in the catalogue with an indication of their status.
Accessibility	<p>The Artexte metadata elements are mostly consistent with concepts used within the Fine Arts community and have evolved over the years in relation to the evolution of the collection.</p> <p>However, some difficulties arise when users and new staff are confronted with the classification system (COTD) used at Artexte. In practice it allows users to find the location of an item in the collection by indicating the name of the physical file in which the document resides. However, this classification system only describes the presence of physical files related to artists, art organizations, subject/disciplines and events; it does not provide an intellectual overview of the content of the collection. The relationship between the files classification system and the actual content of the collection can be a source of confusion.</p> <p>For example, a search in the files database for a particular artist (SZILASI, GABOR) would indicate that there is a file called 410 – SZILASI, GABOR. A user consulting this file would find a limited number of documents compared to</p>

	<p>what they would find when doing a general keyword search for the name 'Gabor Szilasi', which would return documents located in a series of different physical files.</p> <p>This system can present an intellectual barrier to access for users, as a result some instruction is usually provided when users visit the library to ensure that they are able to effectively conduct research.</p>
Conformance to expectations	<p>As a result of the way in which bilingual metadata has been handled some discrepancies have arisen in the use of the English and French language subject descriptors. Documents are catalogued in French by default unless the language of publication is English. Descriptors are assigned in the cataloguing language and in both languages if the publication is bilingual. Over time the list of English and French descriptors has diverged and equivalencies between the concepts in each language have become inconsistent. For example, a user searching for the French subject heading 'Couleur' will not find any of the records identified with the English equivalent of 'Colour'. In a bilingual environment researchers often prefer to be able to easily switch back and forth between results in each language or to see results in both languages. These expectations are not met by the current usage of subject metadata.</p>

4. Choice of Open Access, Open Source & EPrints

4.1 Open Access

The 2002 Budapest Open Access Initiative describes open access as the “world-wide electronic distribution of the peer-reviewed journal literature and completely free and unrestricted access to it by all scientists, scholars, teachers, students, and other curious minds” [5]. The benefits of open access include: user access to relevant literature, increased visibility for authors and works, increased readership and impact, overall lower costs than traditional forms of dissemination [5]. Indeed, the majority of studies published on the so called “OA advantage” provide evidence that open access articles are cited more often [4, 12, 20], although the literature also includes empirical studies where the OA citation advantage is not replicated [8, 9]. The Budapest Open Access Initiative makes it clear that open access should apply to material that “scholars give to the world without expectation of payment”, and that there are degrees of access permissions: read, download, copy, distribute, print, search, or link to the full texts, crawl for indexing, pass as data to software, and any other lawful use [5]. Academics and research funders have expressed a commitment to open access by signing various petitions, declarations and statements in support of open access [25, 26]. Although notable support for open access has come from the medical and scientific funders and researchers, creative disciplines are also interested in the opportunities afforded by this model. Particularly relevant to the Fine Arts is the 2008 International Association of Research Institutes in the History of Art (RIHA) Resolution on Copyright, cautioning that “a regime which is unduly protective of the interest of existing rights holders” can stifle the advance of creative and scholarly work: “neither copyright nor licensing

rules should inhibit the development and diffusion of original scholarly research” [16]. The RIHA resolutions recommend that copyright holders including publishers, galleries, museums, and collecting societies use broad and effective copyright exemptions for purposes of research, private study, criticism and review. It recommends that the definition of ‘research’ encompass analysis and publication, and defines it to be non-commercial when funded by a research council or charity [16]. In the case of “commercial research” outputs, RIHA recommends that any charges for its use for further research and private study be reasonable [16].

Artexte is interested in the potential for an open access platform to provide a subject-specific repository for a community of researchers, writers and artists in the visual and media arts. In this way, the principles of the open access model and its benefits could be applied outside of academia to a discipline that often functions with similar means of public funding and goals in communication of findings. An essential component of the technological framework in supporting this model of dissemination is the ability for scholars and publishers to self-archive their published research findings, data sets, and multimedia artefacts. The self-archiving function provides autonomy to creators while the technological framework provides increased visibility for their work.

Individual researchers self-archive academic articles to create an additional point of access for material that is usually also available through toll access on academic publishers’ web portals. In the case of Artexte, however, the art documentation has up-to-now often only been preserved in paper format. Thus, the new service to be provided by Artexte’s digital repository consists of digital self-archiving of content that is not digitally available elsewhere. The distinction between self-archiving and open-access publishing becomes blurred in the case of Artexte’s holdings since the depositing users will consist of the artists, scholars and writers as well as the publishing organizations such as museums, galleries, artist-run centres in the visual arts community. Artexte digital repository will function both as an open access dissemination platform for organizations as well as a means of self-archiving for independent artists and researchers.

4.2 *Choice of Open Source and EPrints*

The Registry of Open Access Repositories (ROAR) [34] identifies the digital repository software with the most installations to be DSpace and EPrints, both are free/open source software. The advantages of open source software include: low initial cost, accessibility to evaluation without a limited trial period, availability to develop software enhancements without the need to convince a corporation of the enhancement’s financial viability [6]. Moriso et al.’s study of commercial off-the-shelf software (COTS) based development confirms that the dependence on the vendor was one of the major issues in this type of development [23]. Rhyno summarizes the natural synergy between digital libraries and open source software: “Digital libraries and open source software are a natural outgrowth of the open models of exchange that help societies grow and prosper” [28]. Because Artexte is a small organization with limited resources and a non-profit public mandate to promote research, we chose to evaluate available open source repository software for our project.

The usage scenario for the digital repository software is that of a new platform for processes that are essential to the mandate of Artexte. Cruz classifies the considerations in selecting open source software for “mission critical” processes into the following categories: functional (e.g.: essential functionality covered), technical (e.g.: target platform support, reliability, maintainability), organizational (e.g.: active maintenance, sufficient support, long life existence), economical (e.g.: sustainability, flexible maintainability according to individual needs) and

political considerations (e.g.: possibility for influencing further development according to individual needs)[7]. Singh et al. [32] identified a list of evaluative criteria specific to open access digital repository software: adoption, maturity, support, installation, system requirements, globalization, platform support, scalability, authentication, access control, metadata standards, plugins and scripts, object format support, database support, storage abstraction, sustainability, interoperability, developer ecosystem, search engine optimization, upgrade, search, performance, and migration. Singh et al. [32] also summarized the common workflows that form the essential functional requirements of digital repository software: consume (e.g.: navigation, browse, search and download of content files), submit (e.g.: signing up as a depositor, inputting metadata into online forms), accept (e.g.: review and accept/reject deposited items, notify depositors regarding their deposit status, editing and resubmitting), batch (e.g.: bulk importing of content into the repository).

The evaluation, comparison and selection of open access digital repository software has been reported on by various organizations [21, 27, 29, 31, 36]. One of the first comparative evaluations was commissioned as a part of the Open Access Repositories in New Zealand (OARINZ), undertaken by a collaboration of Tertiary Institutions lead by the Christchurch Polytechnic Institute of Technology [36]. This report narrowed down the evaluation to three systems: DSpace, Fedora, and EPrints, providing justification for rejecting three other systems for lack of open source compliant database management, extremely complex installation steps, and lack of community support. The strengths of EPrints identified in the OARINZ study were considered highly relevant for Artexte: a large and broadly distributed install base, a number of installations that have augmented the system's baseline capabilities (demonstrating that the system can be readily modified to meet local requirements), low skill level barrier to implement and maintain, a uniform and well documented code base. The final recommendations in the OARINZ report found all three of these systems to be "credible", and so in order to make a recommendation, the authors re-focused on two criteria: 1) a system that can aggregate and offer a federated search across a national network of separate repository systems, and 2) feature rich system with low implementation and support overheads. They recommend Fedora due to criteria (1), but they also conclude that EPrints is "the best candidate for a self-configuring solution for institutions wanting to set up and host their own repository". Artexte's requirements do not include criteria (1), so the OARINZ [36] recommendation that "the less complicated software architecture of EPrints gives it the advantage over DSpace as the self-configuring institutional repository system" is particularly relevant. These findings were consistent with Arthur Sale who narrows down the choice to the same three systems and argues that "World's Best Practice for an institution commencing an institutional repository is EPrints" due to the fact that it is easy to install, maintain and configure [29]. Similarly, Singh & Pandita used open source license and OAI-PMH compliance as a basic criteria for identifying nine candidate software for a self-archiving repository platform in biomedical sciences, narrowed it further to EPrints and DSpace by using the installation base as a criteria, and then selected EPrints for being "comparatively mature and easier to install, customize and maintain" [31]. Singh & Pandita also describe how they were able to augment the software with Medical Subject Headings (MeSH).

In selecting EPrints as our candidate for further detailed analysis of metadata requirements, we focused on the following properties of EPrints:

- All essential functionality is covered (consume, deposit, accept, batch, OAI-PMH compliance, results are accessible to web crawlers for indexing by Google)

- Large install base and maturity of the software is evidence of its reliability, maintainability, long life existence, and product evolution. The install base also includes three other institutions based in the same city as Artexte: Concordia University, Université du Québec à Montréal (UQAM), and Ecole de technologie supérieure (ETS).
- Recognized by others as a system of choice for organizations that aim to launch a credible open source repository system quickly with limited local IT support
- the EPrints plugin architecture with a plethora of community developed additions is evidence for its flexible maintainability according to individual needs and an active community of users
- Technical support options include support from developers of the core EPrints code at University of Southampton, along with hosting options
- The existence of the Kultur project, and the University of Arts London research repository's use of EPrints demonstrated adoption and use within the field of visual and media arts and is evidence of the possibility for influencing development and sustainability.
- Support for multilingual content in EPrints; UQAM developed a French language configuration set for EPrints that they are willing to share with Artexte.

5. Methodology and Result of Mapping Artexte Metadata fields to EPrints

Our working hypothesis is that EPrints would prove to be a suitable match to support the needs of existing Artexte metadata describing the collection. The following sections describe the process undertaken to evaluate and propose a conversion of the Artexte metadata to EPrints. The methodology involved exporting the existing metadata schema and data of Artexte's Monograph and Periodical databases and the conversion, one-by-one, of all of the 49 Artexte fields into those of EPrints schema. We present the metadata that mapped by default followed by the metadata that would map with modification to the EPrints configuration files. We also discuss the configuration and customizations that would be necessary to accommodate Artexte's record set. Table 3 presents a summary of our findings.

In section 5.1, we list the fields that were an exact match. In section 5.2, we describe the fields that would need to be added using configuration files. This is followed by section 5.3, where we describe some of the necessary customizations in detail for fields that require special attention. In sections 5.4 and 5.5, we describe modifications to the Artexte metadata as well as required extension of EPrints so that it can accommodate the multilingual controlled vocabulary of Artexte.

Table 3. Summary of our metadata analysis: Artexte fields categorized into those that map directly, those that can be added with configuration, and those which require special attention.

Artexte fields that map directly to EPrints (15)	Artexte fields to be added to EPrints using configuration files (25)	Artexte fields that require special attention (8)	Artexte fields to be removed during migration (1)
Creation Date	Variant Title	Publisher	Geographic code
Modification Date	Preceding Title	Language (of eprint)	
Periodical Title	Succeeding Title	Author	
Monographic Title	Subscription	Exhibition	
Periodical Volume	Last Subscription	Statement	
Issue Number	Subscription Expiry Date	Artist	
Date of Publication	Subscription notes	Event	
Public Notes	Publication Frequency	Art Organization	
Language	Fonds Holdings	Subject/Discipline	
ISBN/ISSN	Fonds Summary		
ISSN	Call Number (COTD)		
English Abstract	Location		
Principal Subject Identifier	Document type (Primary)		
Related Web Link	Document type (Secondary)		
Internal Notes	Physical Description		
	Cataloguing Level		
	Price		
	Value		
	Donor Name		
	Fiscal Year		
	Source of Evaluation		
	Acquisition Notes		
	Resume (French language abstract)		
	Edition Statement		
	Status (public or private)		

5.1 Metadata elements that mapped by default

The default EPrints data model is that of top-level containers called “eprints” containing 0, 1 or multiple documents. Documents can contain 1 or many files. The system comes with a default set of metadata fields for both eprints and the documents associated with them. Most of the metadata fields are associated at the eprint level, but some are associated with the document, for example: embargo period, language.

Fifteen fields in Artexte mapped by default to EPrints metadata elements (see Table 4). These fields would not require modification or customization in EPrints or Artexte source data. They represent basic bibliographic information: title, language, date, abstract, keywords, ISSN, volume and issue.

The only discrepancy found in the structure of these fields is as follows: EPrints uses two fields in combination: Date and Date Type to express the concept of Publication Date, whereas Artexte employs a single field.

Table 4. Metadata Mapping from Artexte into EPrints – Exact Match.

Artexte Fields	EPrints fields
Creation Date	timestamp
Modification Date	last modified
Periodical Title	publication
Monographic Title	book_title
Periodical Volume	volume
Issue Number	item_issues
Date of Publication	date date_type (published)
Public Notes	note
Language	language (of document)
ISBN/ISSN	isbn
ISSN	issn
English Abstract	abstract
Principal Subject Identifier	keywords
Related Web Link	related_url
Internal Notes	suggestions

5.2 Metadata that mapped with modification through EPrints configuration files

Although many fields mapped directly between EPrints and Artexte, others required modification. 33 fields would need to be added or modified to accommodate Artexte metadata. Fortunately, EPrints is a highly configurable system, allowing for customization of default eprint metadata fields using configuration files such as *eprints_fields.pl*. A portion of these are a result of the repository doubling as a catalogue of journal holdings (variant title, preceding title, succeeding title, periodical holdings, etc.). Other fields are unique to Artexte's subject specialization in art documentation (e.g. classification fields for artist, curator, organization, event exhibition statement.) Table 5 lists the fields that would need to be added using configuration files.

Table 5. Metadata Fields to be Added Using Configuration Files:

Three fields for keeping track of the history of name changes for journals: Variant Title, Preceding Title, and Succeeding Title
Four fields for periodical subscription information for internal use: Subscription; Last Subscription; Subscription Expiry Date; Subscription notes
A field to explicitly state the Publication Frequency
Two fields for describing Fonds Holdings
An identification number used for Call Numbers (COTD)
An additional field for items located outside of the regular collection
Two fields to describe primary and secondary Document types
A text field to specify Physical Description
An Internal Note about the cataloguing level for the item
Six fields associated with administrative functions: Price, Value , Donor Name, Fiscal Year, Source of Evaluation, Acquisition Notes
In addition to a field for the Abstract in English, an additional French language Abstract is required
A field to record Edition Statements
A Status field to mark records as private or public

One field was removed from the Artexte metadata during the mapping: the Geographic code field, which we found in our metadata quality evaluation to be inconsistent and no longer of use to researchers.

5.3 Metadata that require special attention

This section provides a description of the modifications necessary for four fields which exist by default in EPrints, but would require modification to become compatible with Artexte metadata.

It also discusses four new fields that would be created in EPrints to hold metadata from Artexte’s COTD classification system.

Example modification 1: Addition of custom creator roles

Artexte uses two Author fields (primary and secondary), which can be mapped to the creator field in EPrints. However, at Artexte an author may have multiple roles in relation to an object in the collection. As a result a custom list of roles was developed for authors which reflects this multiplicity. For example, in addition to the roles ‘Artist’ and ‘Author’ there are also roles such as ‘Artist and Author’ and ‘Curator and Author’. The Kultur plugin provides an extended list of contributor roles specific to the domain of Fine Arts which includes possibilities such as ‘Curator’, ‘Director’ and ‘Illustrator’ but these additional multiple roles will need to be added to this controlled list to accommodate Artexte metadata.

Example modification 2: Publisher and Place of Publication is merged into a multiple compound field

The publisher and place of publication information are two separate default fields in EPrints. Since many works catalogued in Artexte are co-published between museums and galleries, it would require configuration to make it possible to store multiple publishers and multiple corresponding places of publication. The three configuration options for this include:

1. Create three separate fields, Publisher1, Publisher2, Publisher3.
2. Flag “publisher” and “place of publication” as “multiple”, rely on proper order of input.
3. Flag “publisher” as multiple and compound field, adding “place of publication” as one of its subfields.

Option 1 has the limitation of accommodating only 3 publishers. Option 2 relies on order of user input to associate place of publication with the publisher, which is error-prone. Therefore, Option 3, the merging of publication with place of publication into a multiple compound field, is the recommended option. Figure 3 shows the multiple compound publisher field as it appears to a depositing user.

Figure 3. Multiple Compound Publisher Field in EPrints (User Input).



This modification is accomplished using eprint_fields.pl configuration file, the two default text fields are merged into one multiple compound field. Table 6 shows the EPrints configuration syntax that is used to accomplish this.

Table 6. Shows Default Eprints Configuration and the Customization that Turns the Publisher into a Multiple Compound Field.

Default eprint_fields.pl	Customized
<pre>{'name' => 'publisher', 'type' => 'text', },</pre>	<pre>{'name' => 'publication_info', 'type' => 'compound', 'multiple' => 1,</pre>

<pre>{'name' => 'place_of_pub', 'type' => 'text', }</pre>	<pre>'fields' => [{'name' => 'publisher', 'type' => 'text'}, {'name' => 'place_of_pub', 'type' => 'text', }], <input_boxes' 2,="" =>="" pre="" }<=""> </input_boxes'></pre>
---	--

Example modification 3: Language is also a property of eprint

To accommodate Artexte metadata, “language” would need to be added as a property of an eprint. By default, EPrints associates language metadata only at the level of uploaded documents. For Artexte, the language of an item is an important piece of metadata that has been catalogued and needs to be migrated even for items that have no corresponding digital document uploaded. We will not be uploading documents for this legacy metadata, and therefore, we need a place to migrate the language information. The field will have to be a multiple field (i.e., allowing for more than one language) because many items in the Artexte collection have been published in multiple languages. New self-archived deposits into the repository will rely on the document-level language metadata entered by the depositors for each uploaded document to automatically populate the eprint-level language field.

Example modification 4: Event Type extended and Exhibition Statement added

In addition to specific fields for event date, location, title and type that exist in EPrints by default, an exhibition statement field needs to be added which will allow for inclusion of a textual statement about the exhibition. Event types would also need to be extended to include art exhibitions. Table 7 shows the relevant default EPrints exhibition related fields and the additional Artexte fields to added.

Table 7. Shows Artexte Exhibition Fields to be Added and the Default EPrints Event Fields

Artexte Field to be added	Default EPrints fields
Event Exhibition Statement event_type “art exhibition”	event_dates event_location event_title event_type: (conference, workshop, other)

Example modification 5: COTD Classification

There is one case where it would be more advantageous to adjust Artexte's cataloguing practices rather than modify the EPrints metadata structure. COTD is a call number system used at Artexte to identify the location of each physical item within the collection. The COTD can refer to an

artists' name, an organizational body, an event, a subject or a discipline. It is constructed using an identification number combined with any of the types of headings mentioned above. In our review of metadata quality, we discovered that the legacy COTD system can present an intellectual barrier to access by users. As discussed by Khoo & Hall, these types of sociotechnical issues in metadata mapping can be the source of significant work, and in this case the mapping of COTD values to EPrints has been a priority [17].

In the context of self-archiving online, constructing these relatively complex COTD numbers is not necessary, because each electronic item has its own unique URI assigned by EPrints. However, for those EPrints that have a corresponding physical document (thus attributed a COTD), it makes sense to keep track of the facets in separate taxonomies instead of trying to combine them into a single string. To remodel the current COTD system, it was decided in effect to 'explode' the existing classification into individual facets that will describe each object in the collection. Four new fields will be created in EPrints to accommodate these taxonomy values: Artist, Art Organization, Event, Subject/Discipline (see Table 8).

EPrints has a native support for taxonomies, which can be used to generate browse screens. Four such taxonomies (artist, art organization, event name, subject/discipline) would be created from the COTD number, and the latter would continue to be assigned internally only for those items that have a physical location at Artexte.

Table 8. COTD Number Components and Corresponding Browse Screens to be Created using EPrints

COTD Number Component	Browse Screen/taxonomy
Artist Name	Artist Name Browse
Art Organization/City Name	Art Organization Name Browse
Event Name/City Name	Event Name Browse
Subject or Discipline	Subject Browse

5.4 Modification to Artexte Metadata

Bilingual Metadata and User Interface

The migration to EPrints provides us with an opportunity to make some changes to the way that bilingual metadata is handled, and as a result to the way that users can interact with content based on language. In the existing Artexte catalogue, search and browse interfaces are bilingual as shown in Figure 4.

Figure 4. Artexte Bibliographic database public web interface

Opérateurs logiques / Boolean Operators :
 & = et / and
 / = ou / or
 ! = sauf / not
 ^ = troncature / truncation

EX.
 abstraction & feminism (abstraction and feminism)
 sculpture / video (sculpture ou vidéo)
 snow ! michael (snow sauf michael)
 organization (organization, organisation)
 technolog* (technology, technologie, technological, etc)

Recherche Détaillée / Detailed Search

Artiste(s) / Artist(s)
 AND [] Ex. wall, jeff :: wall

Auteur(s) / Author(s)
 OR [] Ex. biron, normand

Titre / Title
 AND [] Ex. Sightlines

Editeur(s) / Publisher(s)
 AND [] Ex. Galerie 101

Descripteur(s) / Subject(s)
 AND [] Ex. sculpture, abstraction

Inscrire des termes et appuyez sur le bouton « entrée » pour effectuer la recherche.
 Type your terms and press the "enter" key to initiate the search.

Vous pouvez entrer plusieurs mots clés sur une ligne - l'opération défaut est ET. Les lettres majuscules et les accents ne sont pas nécessaires. Cliquez sur le nom d'un champ pour une liste des termes à chercher

You may enter more than one keyword on a single line - the default operator in this case is AND. There is no need to use capitals or accents. Click on the name of a field for a list of search terms

Limiter votre recherche / Limit your search:

Date(s) de Publication **Type de [Document] Type**

AND [] AND []

<1990 (avant/before 1990)
 >2001 (après/after 2001)
 1990:2001 (entre/between 1990 & 2001 inclusif/inclusive)

Code Géographique **Langage / Language**

AND [] AND []

Ex. CAN-ONT-TORONTO
 Utile pour trouver des publications concernant un région géographique
 Used to find publications relating to a geographic area

For many Canadian organizations, where English and French are the official languages, providing access to bilingual content can be a challenge. There is the language of the interface itself that must be managed, whether there are separate translations for each language or whether the interface itself is bilingual as in the example above (Figure 4). EPrints has the capability of generating end user interfaces in multiple languages through the addition of translation sets to configuration files.

In a Canadian research context it is ideal to provide users the ability to see all relevant results, but then also to filter results by French or English language values. This would mean that users can control the language of search results independently of the interface language. The addition of the new field for publication language at the eprint metadata schema level will accommodate this kind of functionality.

In addition to the interface language, a strategy is required to determine how results will be displayed to users. To date Artexte has stored bilingual COTD values in the same field, for example the file name for the following organization:

311 - CANADA - CONSEIL DE RESSOURCES HUMAINES DU SECTEUR CULTUREL / CULTURAL HUMAN RESOURCES COUNCIL (CRHSC/CHRC)

This is not very effective metadata for use in the Art Organization browse screen in EPrints. In the migration to EPrints we will separate the bilingual values in our metadata into independent French and English values for browsing in each language.

5.5 Custom EPrints development: bilingual thesauri

Both the Document Type field and the system of COTD classification numbers used at Artexte include metadata to describe Scope Notes, Narrower Terms, Broader Terms, Related Terms, Use and Use/For. For example, the COTD

330 - GALERIE HORACE (Sherbrooke) **USE** 330 - SPOROBOLÉ (Sherbrooke)

indicates a name change for this organization. Likewise, the COTD

410 - BILODEAU, ANN **USE** 410 - ONIRICO (Collectif)

indicates that materials related to this artist have been classified under the name of the artist collective that she belongs to.

This metadata has been maintained at Artexte to reflect frequent changes in names and associations in the domain of visual arts, and to ensure that users will be able to find all pertinent content related to their research.

Artexte metadata also includes a thesaurus of document types which would need to be added as a separate descriptive field. The default document types in EPrints include: Article, Book section, Monograph, Conference or workshop item, Book, Thesis, Patent, Artefact, Show/exhibition, Composition (musical), Image, Video, Audio, Dataset, Experiment, Teaching Resource, Other. The Artexte document types are listed in Table 9 below.

Table 9. Artexte Document Types

Journal	Brochures, booklets, leaflets, pamphlets
Diaries	Conference, colloquium, symposium publications
Comic strips, graphic novels	Proceedings, transcripts
Novels	Dictionaries, encyclopedias, lexicons
Poetry	Audio, visual, multimedia
Calendars	Artists' books
Correspondence	Mail art
Directories	Multiples
Monographs	Activities reports
Off-print	Annual reports
Scripts, theatre plays	Research reports
Government documents and publications	Statistical reports
Facsimiles	Flyers, posters
Guides and manuals	Work of art
Interviews	Cards, deck of cards, postcards
Annuals	Invitations
Databases	

Professional literature	Auction catalogues
Periodical issue	Catalogue
Insert	Distribution catalogues
Audio recording	Event catalogues
Computer artwork	Exhibition catalogues
Video recording	Anthologies, essays, collections
Zine	Bibliographies
Programme	
Boxed set, collection	
Exhibition ephemera	
Glossaries	

For example, the term “Video recording”, has the following Use reference: “Audio, visual, multimedia”. This allows the Artexte system to point the cataloguer to the appropriate controlled term. If the cataloguer browses to “Video recording”, for example, they would be prompted by a USE reference to choose the preferred term “Audio, visual, multimedia”. The Document type thesaurus is bilingual and cataloguing is done in the language of publication. The French term “Enregistrement vidéo” has a corresponding USE reference to “Audio, visuel, multimédia”.

Although EPrints does have built-in capacity to store taxonomies with hierarchical relationships and language represented; adding the associative relationship (RT), scope notes, and Use/UF needed for the above Classification and Document type fields would require a significant software development effort. Maintaining relationships between equivalent English and French terms would also require changes to the underlying EPrints taxonomy schema. The EPrints data model would have to be extended to include the additional fields, and corresponding functionality would have to be developed to generate the appropriate links on the user browse and deposit screens that are based on a thesaurus. The web administration interfaces that allow librarians to modify and update controlled vocabularies would have to be extended to include the additional fields. The additions necessary to accomplish this modification demonstrate a software development effort that represents a branch or an entirely new plugin to EPrints, as opposed to a relatively simple change to configuration/customization scripts.

As a useful reference point for this type of customization the Kultur project metadata report identified the following “Art/Design Categories”: Other, Animation, Architecture, Audio Work, Book art/Artist’s book, Ceramics, Conservation work, Craft, Curation, 2D Design, 3D Design, Digital art, Drawing, Exhibition/show, Fashion, Film, Illustration, Installation, Painting, Performance, Photography, Printmaking, Public art, Sculpture, Site-specific work, Sound art, Teaching, Textiles, Theatre, Typography, Video. However, the situation at Artexte requires many other Document type terms because the cataloguing that is done there is at the level of all supporting materials, including exhibition catalogues and pamphlets as well as monographs and journal articles about the art works.

4. Conclusion

This paper includes a descriptive analysis of a collection of metadata on research, interpretation and dissemination activities in the visual arts in Canada that has been developed for a user community of artists, art historians, curators and art enthusiasts over the last 30 years. We described the current database structure of Artexte, its controlled vocabulary, and presented the results of a manual evaluation of metadata quality. We believe this descriptive analysis to be a valuable contribution as a case study of digital library structure in Fine Arts documentation. Descriptive analysis of metadata collections that have been developed to serve user communities are a necessary step towards progress in solving challenges of interoperability of metadata on the web and the usability of digital library software.

In this study, we analyzed the compatibility of this metadata collection with leading digital repository software used for open access, EPrints. We discussed the open access model, its advantages and relevance to Artexte and conclude that it is a suitable model for adoption by the organization. We also discussed the rationale for selecting an open source system, EPrints, for the comparative analysis. In our comparative analysis, we listed the metadata elements that mapped by default, the metadata elements that mapped with modification through configuration files and described the structures that require special attention. The language field is particularly interesting in that it challenges the EPrints data model that assigns this property to uploaded documents, not the eprints that “contain” the documents and the majority of the metadata for them. Lastly, we identified an area that would require custom development effort in EPrints: accommodating some of the controlled vocabulary in Artexte (e.g., document types, organization names, subjects) demands an extension of the EPrints taxonomy model with thesaurus semantic relationships (scope notes, related terms, use and use/for). Although thesaurus functionality in digital repository software does seem to be an active research area (e.g.: [22]), we did not find it in any digital repository software comparison and/or evaluation documentation.

We conclude that the EPrints platform would be suitable to the needs of Artexte's collection and bibliographic data management. This conclusion is supported by the fact that in our comparison of Artexte metadata to the EPrints metadata schema, we found that of a total of 49 fields, 1 field will be removed during migration, 15 fields mapped by default without need for modification, 25 fields would need to be added to EPrints configuration files using a simple process and 8 fields require special attention. One area that merits further study includes a special render functionality that would have to be written for some of the new fields, such as the new complex publisher information and the thesaurus-based document-type data input and display. The new context of self-archiving, as opposed to professional cataloguing also raises questions of data quality. At this time, we did not determine which fields should be mandatory in the self-archiving context – this could be the subject of future work.

We hope that our description and analysis will prove useful to organizations working with discipline-specific documentation formats that do not always conform to the standards of an academic repository. This includes organizations with similar metadata, metadata aggregating organizations, as well as to open repository platform developers working with research in the creative disciplines.

References:

1. Averkamp S, Lee J (2009) Repurposing ProQuest Metadata for Batch Ingesting ETDs into an Institutional Repository. code4lib journal <http://journal.code4lib.org/articles/1647>. Accessed 03 June 2010.
2. Bradley J, Johnstone L (1994) "Introduction". Sightlines: Reading Contemporary Canadian Art. Montreal: Artextes editions, 1994.
3. Bruce TR, Hillman D (2004) Metadata in Practice, Chap. The Continuum of Metadata Quality: Defining, Expressing, Exploiting. pp. 238-256. ALA Editions, Chicago.
4. Brody T, Harnad S, Carr L. (2006) Earlier Web Usage Statistics as Predictors of Later Citation Impact. Journal of the American Association for Information Science and Technology 57(8):1060-1072. DOI: 10.1002/asi.20373
5. Budapest Open Access Initiative (2002). <http://www.soros.org/openaccess/read.shtml>. Accessed 15 August 2011
6. Corrado EM (2005) The Importance of Open Access, Open Source, and Open Standards for Libraries. Issues in Science and Technology Librarianship. Spring 2005. <http://www.library.ucsb.edu/istl/05-spring/article2.html>. Accessed 8 August 2011
7. Cruz, D (2006) Evaluation criteria for free/open source software products based on project analysis. Software Process Improvement and Practice 11(2): 107 DOI: 10.1002/spip.257
8. Davis PM (2010) Does Open Access Lead to Increased Readership and Citations? A Randomized Controlled Trial of Articles Published in APS Journals. The Physiologist 53(6):197-201. http://www.the-aps.org/publications/tphys/2010html/December/open_access.htm . Accessed 11 August 2011
9. Davis PM, Lewenstein BV, Simon DH, Booth JG, Connolly MJL (2008) Open access publishing, article downloads and citations: randomised trial. BMJ 337: a568. DOI:[10.1136/bmj.a568](https://doi.org/10.1136/bmj.a568)
10. Fay E (2010) Repository Software Comparison: Building Digital Library Infrastructure at LSE. Ariadne <http://www.ariadne.ac.uk/issue64/fay/>. Accessed 8 August 2011

11. Floyd, R (2009) Automated Electronic Thesis and Dissertations Ingest. <https://wiki.dlib.indiana.edu/display/IUSW/Automated+Electronic+Thesis+and+Dissertations+Ingest> (). Accessed 26 May 2011.
12. Gentil-Beccot A, Mele S, Brooks TC (2009) Citing and reading behaviours in high-energy physics. *Scientometrics*. DOI: 10.1007/s11192-009-0111-1
13. Gonçalves MA, Fox EA, Watson LT (2008) Towards a digital library theory: a formal digital library ontology. *Int J Digit Libr* 8:91-114 DOI: 10.1007/s00799-008-0033-1
14. Gray A (2009) Institutional Repositories for Creative and Applied Arts Research: The Kultur Project. *Ariadne* 60 <http://www.ariadne.ac.uk/issue60/gray/>. Accessed 03 June 2010.
15. IFLA Study Group on the Functional Requirements for Bibliographic Records (2009) Functional Requirements for Bibliographic Records. International Federation of Library Associations and Institutions http://www.ifla.org/files/cataloguing/frbr/frbr_2008.pdf. Accessed 11 August 2011
16. International Association of Research Institutes in the History of Art (2008) RIHA Resolution on Copyright. <http://www.riha-institutes.org/resolutioncopyright.html>. Accessed 15 August 2011.
17. Khoo M, Hall C (2010) Merging metadata: a sociotechnical study of crosswalking and interoperability. *JCDL '10 Proceedings of the 10th annual joint conference on Digital libraries*. Gold Coast, Queensland, Australia. June 21–25, 2010. DOI: 10.1145/1816123.1816180
18. Latour J, Tayler F (2009) Contemporary glocal: Positioning a national mandate within a global context. *Art Libraries Journal* 34(1):40-45.
19. Léger D (2006) Le Centre d'information Arttexte : un mandat, et un parcours, atypiques. *Essays in the History of Art Librarianship in Canada*. ARLIS Canada. <http://www.arliscanada.ca/hal/>. Accessed 03 June 2010.
20. Lawrence S (2001) Free online availability substantially increases a paper's impact. *Nature* 411(521). DOI:10.1038/35079151
21. Marill JL, Luczak EC (2009) Evaluation of Digital Repository Software at the National Library of Medicine. *D-Lib Magazine* 15(5/6) <http://www.dlib.org/dlib/may09/marill/05marill.html>. Accessed 9 August 2011
22. McKay D, Shukla P, Hunt R, Cunningham SJ (2004) Enhanced browsing in digital libraries: three new approaches to browsing in Greenstone. *Int J Digit Libr* 4:283–297. DOI: 10.1007/s00799-004-0088-6

23. Morisio M, Seaman CB, Basili VR, Parra AT, Kraft SE, Condon SE (2002) COTS-based software development: Processes and open issues. The Journal of Systems and Software 61:189–199 DOI:10.1016/S0164-1212(01)00147-9
24. Ochoa X, Duval E (2009) Automatic evaluation of metadata quality in digital repositories. Int J Digit Libr 10:67-91 DOI: 10.1007/s00799-009-0054-4
25. Open Access Directory (2011) Declarations in support of OA. [http://oad.simmons.edu/oadwiki/Declarations in support of OA](http://oad.simmons.edu/oadwiki/Declarations_in_support_of_OA). Accessed 11 August 2011.
26. Open Access Directory (2011) Statements by learned societies and professional associations [http://oad.simmons.edu/oadwiki/Statements by learned societies and professional associations](http://oad.simmons.edu/oadwiki/Statements_by_learned_societies_and_professional_associations). Accessed 11 August 2011.
27. Repository Support Project (2010) Repository Software Survey (November 2010). <http://www.rsp.ac.uk/documents/Repository-Software-Survey-2010-11.pdf>. Accessed 4 August 2011
28. Rhyno, A (2004) Using Open Source Systems for Digital Libraries. Westport, Conn, London : Libraries Unlimited.
29. Sale A (2005) The Key Things to Know. University of Tasmania Eprints Repository. http://eprints.utas.edu.au/223/1/NZ_Workshop_PDF.pdf. Accessed 5 August 2011
30. Sheppard V (2009) Metadata Report. JISC. <http://kultur.EPrints.org/Metadata%20report%20Final.pdf>. Accessed 03 June 2010.
31. Singh S, Pandita N (2005) Building the open access self-archiving repository for the bio-medical sciences at National Informatics Centre. In: National Convention of Medical Library Association of India, 07-09 Nov 2005, Bangalore, India. <http://openmed.nic.in/1108/>. Accessed 11 August 2011
32. Singh SK, Witt M, Dorothea S (2010) A Comparative Analysis of Institutional Repository Software. Fifth International Conference on Open Repositories. Madrid, Spain, July 2010. <http://or2010.fecyt.es/Resources/documentos/GSabstracts/AComparativeAnalysisInstitutionalRepositorySoftware.pdf>. Accessed 4 August 2011
33. University of Southampton (2011) EPrints - Digital Repository Software <http://www.eprints.org/software>. Accessed 3 June 2010

34. University of Southampton (2011) Registry of Open Access Repositories (Browse by Repository Software).
<http://roar.eprints.org/view/software/>. Accessed 10 August 2011
35. Walsh MP (2010) Batch Loading Collections into DSpace: Using Perl Scripts for Automation and Quality Control. *Information Technology and Libraries* 29(3): 117-127
http://www.ala.org/ala/mgrps/divs/lita/ital/292010/2903sep/walsh_pdf.cfm. Accessed 26 June 2011.
36. Wyles, R (2006) Technical Evaluation of selected Open Source Repository Solutions. Open Access Repositories in New Zealand (OARINZ) project, Christchurch Polytechnic Institute of Technology, v1.3, 13 September 2006
http://dlynx.rhodes.edu/jspui/bitstream/10267/30/1/Repository_Evaluation_Document_NZ.pdf. Accessed 9 August 2011