Metadata for Photographs: from Digital Library to Multimedia Application

Anne-Marie Vercoustre[1] and François Paradis[2]

[1] INRIA Rocquencourt,
BP 105, 78153 Le Chesnay, France
Anne-Marie.Vercoustre@inria.fr
http://www.inria.fr/

[2] CSIRO Mathematical and Information Science,
723 Swanston St., Carlton, VIC 3053, Australia
Francois.Paradis@cmis.csiro.au
http://www.cmis.csiro.au/

Abstract. This paper describes the production of an educational multimedia CD-ROM about French rural houses and farms, and how to renovate them without losing their traditional features. The educational message is illustrated with many photographs of non-renovated or renovated houses, and made explicit through comments and descriptions associated with the photos. The paper focuses on the XML metadata describing the photos and the use of this metadata for the automatic generation of Web pages. We first report on the usability of the Dublin Core for interoperable photographs metadata, together with more detailed XML descriptions to support a specific multimedia application. We then show how to generate the Web pages by defining HTML document *prescriptions* which embed queries to the XML metadata, using Norfolk, a virtual document generator. The approach can be used in various applications ranging from personal virtual photo albums to complex virtual museum.

1 Introduction

Digital Libraries provide access to an increasing variety of digital multimedia information, including images, photos, sound and video. For many years Libraries and Digital Libraries have been relying on classification and indexing schemas for supporting retrieval of bibliographic references and documents, while Web search engines use a mix of full text indexing and classification. Searching for non textual material like photos, video and sounds[1] requires indexing through related textual descriptions, or *metadata* generally in an external document, but possibly encoded within multimedia formats such as JPEG. Metadata, or information about data, have motivated a lot of international interest and effort

[1] Searching by similarity with a given photos can be done using internal low level characteristics of the photo. We are not addressing this kind of search here.

in the last three years, through the Dublin Core Workshop series [5], or the Resource Description Framework (RDF) defined by the W3C.

The Dublin Core has attracted the attention of formal resource description communities such as museums, libraries, government agencies, and commercial organisations, as illustrated by the many projects listed on the DC site. However there is very little report on the use of DC metadata for photographs. Before the availability of digital photographs, some effort has been spent on cataloguing photographs although they are not very specific cataloguing tools for photographic material (AAT Art and Architecture Thesaurus, LC thesaurus for Graphic Material), nor appropriate standards. As speculated in [9], archivists for photos and graphic collections may have considered their collections as unique and assumed that no imposed standard could work on their collection. This advocates once more for an infrastructure such as RDF that supports the coexistence of complementary, independently maintained metadata packages.

Metadata have been primarily designed to support the retrieval and display of resources. Digital Libraries of the future will not just display raw material stored in their databases, nor will they deliver it in a one-for-all presentation. As claimed in [13], "Searching is not enough". Information will be accessed through user queries, browsing, guided tours, temporal exhibitions, educational courses, etc. The same resource will be used in different contexts, associated with other relevant material. Photographs, for example, will be pulled out with other documents which are associated with them either statically or dynamically. All this information needs to be ready for use during frequent context shifts. The material collected for a particular task needs to be pulled together and kept for near or long-term use in new information artefacts. Users as well as librarians and curators need facilities to construct information compounds or to build flexible presentations of Digital Libraries resources.

This paper presents an example of a simple, yet significant, multimedia application (ultimately a CD-ROM) that is automatically generated from XML Dublin Core metadata describing photos. The DC *Description* element has been enriched with a specific schema to support the target application. A more complex application may require to separate Dublin Core metadata from other XML descriptions and complex documents, but the principle would be the same: the application is generated from page *prescriptions* which define the page model and the queries to retrieve the appropriate elements (photos, texts) from the XML metadata. Links to other pages can also be modelled, and result in an hypertext which is coherent and easy to modify.

The rest of the paper is organised as follows. First we present our application: a CD-ROM about rural housing. We then discuss two problems associated with the production of this CD-ROM: how to express metadata about photographs using Dublin Core, and how to automate the generation of pages with virtual documents using the Norfolk system. The last section will outline some benefits of the approach and will draw some conclusions and further work.

2 The application

We set out to produce a CD-ROM on rural habitations in a small area near Paris, France, where renovation and modern housing have been developing rapidly in the last thirty years, endangering the traditional aspect of French villages to be completely ruined. The aim of the CD-ROM is to make people more aware of this rural cultural heritage and to provide guidance for renovating houses while preserving their traditional features.

About 300 photographs from the various village had been taken in the area to illustrate and support the discourse. They had been assembled into a photo exhibition which toured in several villages during 1992-1993. Following the success of this exhibition we thought it was worthwhile to make this material available on a more permanent and accessible support such as a CD-ROM. Like the exhibition, the multimedia presentation will use mostly photos for supporting the educative message through illustrative and comparative examples. About 150 photos have already been scanned for use in the multimedia presentation. The main advantage of an hypermedia presentation over a traditional exhibition is that some photos can be reused several times, in different contexts, for illustrating different aspects of rural buildings. For example a photo of a nice and well renovated house could be used as an example for roof, windows or facade renovation, associated with the relevant and focused comment.

The CD-ROM, as a more perennial support, needs to provide also detailed information about the buildings, so that the photos may become part of the village archives. Overall information will include:

- factual information (full address, date of the photograph, state of the building, etc.)
- comments and description about the main interesting features
- comments and appreciation about the quality of the renovation (this is somehow tricky to avoid litigation).
- comparisons of the same building over time (when photographs or postcards are available)

The user can navigate the CD-ROM using a Web browser. Figure 1. shows the navigational structure. Navigation can be guided by village/street, theme (roof, windows, etc.), or (not shown in the figure) by examples of renovation (good ones as well as bad ones, on different aspects) or comparison before and after renovation (when available). The description page for a photograph can depend on its context: for example, the description will be different if the user is looking at "roofs" or "windows".
Examples of photos can be seen in Figure 2.

The digital photos have been processed by a photo laboratory from the colour negatives and put on a read-only CD-ROM. We thought initially of using the KODAK-CD standard to keep the best definition of the photos as recommended in [8]. A KODAK Photo-CD can store a maximum of 660Mb of photos (about 100 photos). As we intend to store at least 150 photos, and ultimately much

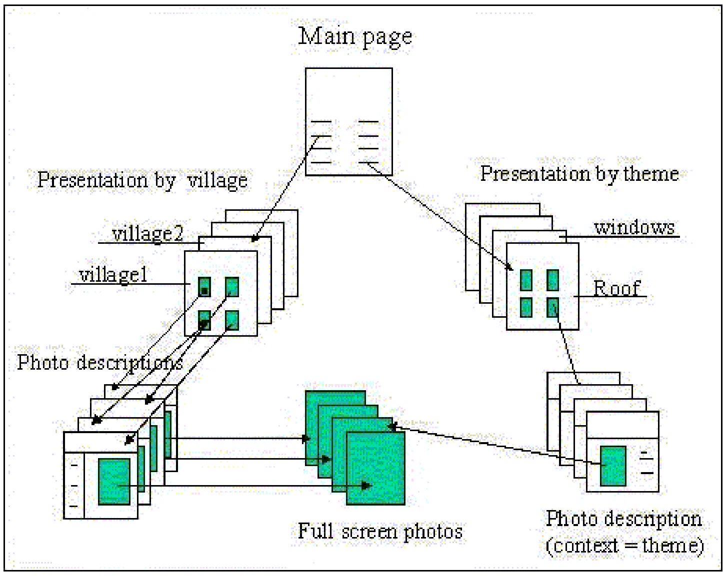

Fig. 1. Structure of the CD-ROM

Fig. 2. Example of photos for house fronts

more, it would have been too costly for the budget of the local association who is producing the final CD-ROM(s). Since the application was intended mostly for public education, *working images* were sufficient for our purpose. We therefore decided to store digital photos as 24-bit images compressed in JPEG, of dimension of 800x537 pixels. It results in image files of size ranging between 42k and 300k, which would allows us to store millions of images on one photo CD-ROM. From these images we produced thumbnail images (200x 134), also called *browsing images* in [8], by reducing the images by 25%, which results in image files of size between 8 and 35k. If smaller images were necessary for an application (iconic images) there will be produced at display time by setting the attributes in the HTML IMG tags.

3 Metadata for photographs

Although photographs may be more explicit than a long discourse for an humans, they don't describe themselves in term of content as texts do. For texts, authors use many clues to indicate what they are talking about: titles, abstract, keywords, etc. which may be used for automatic cataloguing.

Searching for photos must rely on manual cataloguing, or relate texts and documents that come with the photos. Creating standard metadata for photograph content is difficult since one photo may have very different meanings, depending on the context, of whether it is part of a collection, or even on the purpose of a particular end user. Our collection was very specific, and there was no standard for describing this kind of buildings. At least the Dublin Core offers a framework for describing digital resources and we wanted to experience how it would work for our photos. Our main motivation for using a standard like Dublin Core was to make our photos reusable for other experiments or applications, by ourselves or by other groups.

For example one could think of creating a CD-ROM by village and compare how a house looks today with how it looked in 1900 by comparison with the many available postcards.

3.1 Dublin Core

The Dublin Core metadata set [5] is intended to promote and develop the metadata elements required to facilitate the discovery of resources (documents and images) in a networked environment such as the Internet and support interoperability amongst heterogeneous metadata systems. Dublin Core metadata set consists of fifteen descriptive elements. Commonly understood semantics for these element is described in the reference description [7] which has been largely stable since 1996, and known as DC 1.0. Although the first aim of Dublin Core is to offer a minimal set of core elements to achieve interoperability, extensibility has been an important issue since the second DC workshop. Early experience with Dublin Core deployment has made clear the need to support qualification of elements for some applications. Thus, a Dublin Core element may be expressed without qualification or with qualifiers that refine its semantics. Three qualifiers were defined: sub-elements, language and scheme. A *sub-element* specifies a facet of a given element. The intent is to narrow the semantics of a field, not to extend it. The dot-encoding convention used for this qualifier reflects this. The *language* qualifier indicates the language used for the content of the element. Finally a *scheme* qualifier specifies a context for the interpretation of a given element. Typically this will be a reference to an externally-defined scheme or accepted standard. Dublin Core defines *scheme* as a qualifier that "provides a processing hint that may be used by an application or a person to make better use of the element that is qualified".

We expressed DC metadata in XML as shown in the example given in Appendix 1. A more formal description using RDF and following the guidance on expressing Dublin Core within RDF [12] can be read in Appendix 2. In the next section we describe in detail our use of Dublin Core and the few qualifiers we added (sub-elements and schemes). Our schemes are XML schemas and values for the corresponding elements are therefore expressed also in XML.

3.2 Report on our use of DC

Although our collection of photographs is rather modest in size, our use of DC has given rise to a number of small difficulties which are worth reporting on. As mentioned before, very little report on using DC for images and photographs is available today.

We describe below for every DC element, its use in our application.

Title: What is the title of a photo is not very well defined. It could be seen as the caption of an image, but it is generally dependant of the context. Think of the kind of legends you may write in your family photo album. It may be very informative, or just humorous. In our case we put a very general description that may be used in various contexts, and of some use for general retrieval.

Author: The person who has taken the photograph.

Contributor: The contributor could be the lab or person who developed the photo, or scanned it. Each contributor could have a role specified with a sub-element qualifier. Here we just mentioned the photo shop who scanned the photos.

Subject: It should be a list of keywords, in a standard terminology for buildings and architecture. Since rural habitat has not yet attracted much interest as cultural heritage, there is no classification schema specifically relevant to these buildings. So we use common words (i.e. windows, roof, facade). Subject element is more intended for interoperability with other applications and we are not using it to query the metadata. Instead we use the sub-elements of the Description element since the photo will be retrieved only if there is a specific description for this subject.

Description: Each photo has two descriptions: a general one usable in the presentation by village, and a more detailed using a scheme qualifier specific to the application. Our schema contains XML elements for the parts of the house that are being described (roof, roof windows, windows, doors, coating, front, etc..). It is also used to describe the general state of the building: renovated, poor condition, etc.

Publisher: The publisher of the CD-ROM. There is a problem here if we want to make the material reusable: the publisher would depend on the specific application the photo is used in. In our case it was not clear what to put in. We chose to mention the Association who supports the creation of the CD-ROM, although the photos (and their metadata) could be reused in another context.

Date: The date when the photograph was taken (not scanned). At this point it seems that Dublin Core requires to use Date for the date of the digital resource, and possibly source.date for the date of the original photo. For us it was typical of the problem that arises with photos. We would rather consider the photo as the resource, and the various digital images as surrogates for this photo. We will come back to this point with Format. For our application what is important is the date of the original photo which marks the state of the building at a determined period.

Type: The type of the resource which is just an "image" in term of the DC recommended Type categories. We also mentionned whether it was a colour or black and white photo, in a second type element.

Type.quality: The quality of the photo. There is no specific element in DC to speak about the intrinsic quality of an image or a photo. Although it may be somehow subjective, it is an important information when the user or the application want to make choice between equivalent photos to display. Puting it as a sub-element is certainely not conform to DC requirment for qualifier, but there was no appropriate place to add this information, using Dublin Core.

Format: The format of the digitised image (MIME/JPG).

Format.orientation: Whether the photo must be seen horizontally, or vertically. This is an important information when selecting photos for multimedia presentation, since vertical photos can be grouped differently as horizontal ones to fit in an usual screen size. **Format.full** and **Format.small** are referering to the existence of two surrogate formats for the photos. Again there was a difficulty here in describing various format/size for the same photo.

Identifier: a reference to the actual photo. We could not use an URL since the photos will move to a CD-ROM, with a different hierarchy. We use an URI that reflects the relative organisation of the application.

Source: A reference to the original photo, or its negative; the value of the reference will strongly depend on where the original is archived, whether it is a personal collection or an Archive Department. A specific SCHEME could be useful here.

Language: Not applicable for photographs.

Relation: Relation to the overall collection or to another photo for comparison. Relation could in principle be used for reference to surrogates of the same photo, but we did not like this solution which will make seaching for a photo a very complex process.

Coverage: The date and location of the object in the photo (i.e the building). **Coverage.date**: The date of the object being photographed We did not use this element since we had little information on the construction date of the buildings. *Coverage.location*: The precise address of the building since the photo could serve as a future archive for the villages. We introduced an address scheme (number, street, town, country), to be able to retrieve and sort out the photos by village, street, or house.

Rights: The copyrights for the photograph. There is some ambiguity between the rights for the source and for the digitised photograph. We mean here the source version, and by extension any digitised version of the photograph. Copyrights may be much more complex.

In resume the difficulty with metadata for photographs is to make the distinction between the information about the photo and the information about what is represented by the photo. Retrieval of the photo will be mostly done from the latter while rendering of the photo has nothing to do with its content. The existence of various "versions" or digital surrogates from the original photo makes it more confused, as reflected in the use of the Format element. These problems have been already identified in [18].

3.3 Metadata for describing the scanning process

[2] makes strong requirements on what information should be kept on the format of a digitised image and on the scanning process. This information is mostly irrecoverable if not recorded at the time of the capture. In large archiving projects this would be a major requirement, while less critical for casual user or small applications such a CD-ROM. These information may be important to be able to display long term archive images in many years, or to create large photograph workbenches. In our case we did not store any information about the scanning process other than the name of the laboratory who did the work. We feel it was not necessary in regard to the type of photos and precision of the digitalised images.

3.4 Application specific metadata

As said before, metadata specific to the application were introduced as extra tags in the various *Description* elements, using a Scheme qualifier. For example, the description of a "window" is expressed as follows:

```
<DC:Description SCHEME ="habitat">
      <rural> <window example="good">
        The windows are heigther than large (1/3), with 6 glasses.
        Note the irregularity of the window disposal, with a large
        plain area on the right.
```

```
            </window>
        </rural>
    </DC:Description>
```

For more complex applications this information could be stored in a separated XML document, with a possibly complex DTD. For simplicity, we choosed to store it in repeated Description elements.

4 Virtual Documents

Our CD-ROM will have a large number of pages, even for only 150 photographs since each photo may correspond to several presentations depending on the context where it is displayed. Moreover we intend to scan and add more photos in the future and design more personalised versions of the CD-ROM for didactic usage in a specific village. Clearly we need a tool to automate the creation of pages, which would allow to select photos and group them by themes, villages, etc.

Our solution is to generate pages as *virtual documents*, i.e. to write HTML "templates" with normal tags and text and with instructions on how to generate dynamic information extracted from the photograph collection. In this section we introduce **Norfolk**, our virtual document generator, give some examples of document prescriptions to generate our CD-ROM, and describe the set of prescriptions. We conclude with a discussion about dynamic generation.

4.1 The Norfolk system

Norfolk is a system for the generation of virtual documents. Initially developed to promote the idea of *information reuse*, it has been used for Web delivery (see [14]) to see how it is used to generate and maintain the TED Web Site (at http://www.cmis.csiro.au/TIM/), and more recently for customised information delivery, i.e. how to select, modify or generate the information to satisfy a particular user need.

In Norfolk, the instructions to generate a virtual document have a *descriptive* rather than a *programmatic* style, and, more importantly, a common data model is defined to view information coming from a variety of sources as a document-like structure. Information sources can still be queried in their native format (e.g. SQL for a relational database), but results are then "translated" to a tree structure similar to a document parse tree. A query language allows selection or modification of the tree, and supports useful operations for document structures. For example it supports extraction of *semi-structured* information, i.e. information whose exact structure is not known or can vary. For more information on Norfolk data model and language, see [16]. An online demonstration is also available at http://www.ted.cmis.csiro.au/RIO.

Figure 3 shows how Norfolk is used to generate a set of pages. The instructions to generate the virtual documents are stored in *document prescriptions*. The em interpreter takes a document prescription as input, queries data sources such as databases or other pages, and produces a virtual document as output. The "children" or document prescriptions referenced in the virtual document, are then recursively generated. In other words, the generator "follows" the links to document prescriptions in order to generate a network of virtual documents. Cycles and naming for the virtual documents are automatically dealt with.

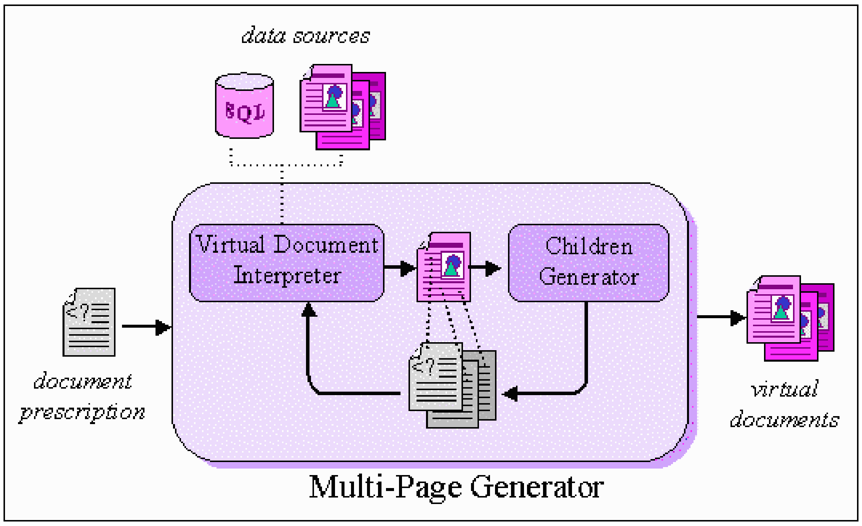

Fig. 3. Generation of pages using Norfolk

4.2 Examples of prescriptions

The following example shows part of a prescription for collecting all the photographs related to one village (passed as a parameter) and to display the list of those photographs, with the address of the places and links to the actual photographs.

The document prescription is a valid HTML document, to which Norfolk instructions are added in the form of process instructions (i.e. <? ... >). The "Part 1" finds the photographs for the village. It first gets the list of all metadata files (an index file, `DIR.xml`, which is dynamically generated, has an <`f`> element for every file) and then iterates through the list, loading the file and checking if it is in the right village. In practice, for efficiency reasons, the list of photographs for every village could be built beforehand, and included in `DIR.xml`. "Part 2" then outputs the photographs in a list, with the address and a link to the image for every photograph.

```
<!-- Takes one parameter: the village name --></font>
<?param "village" as $village>

<!-- Output title (village name) --></font>
<H1>    <?$village>    </H1>

<!--=== PART 1: find photos for village ===-->
<!-- Assign all filenames to $photos --></font>
<?define $photos as url("DIR.xml")..f>

 <!-- build $photosVillage (photos for a village)
      by iterating through the list of all photos --></font>
<?map $xx in $photos>
  <?begin>
  <?define $a_photo as url(str("photos-XML/",$xx))>
  <?if $a_photo..town contains $village>
    <?begin>
    <?define $photosVillage as $photosVillage adopt $a_photo>
    <?end>
  <?end>

  <!--=== PART 2: output list of photos ===--></font>
<ul>
<?map $photo in $photosVillage>
<li>    <?$photo..Address.number;$photo..Address.street>
<A>    <?attributes attr("href",str("../",$photo..DC:IDENTIFIER))>
  <?$photo..DC:SOURCE>    </A>
</li>
</ul>
```

Fig. 4. Document prescription for a Village

The following prescription produces a list of four villages, with a link to their respective page. The `href` attribute, created with the `attr` command, is a link to the village prescription. This link consists of the prescription name, `village.rhtml`, and the village name: for example, the `urlstr` command will produce `village.rhtml?village=Chavenay`). The `attributes` instruction is then used to attach the `href` attribute to the preceding tag, <A>.</A>

```
<html>
<!-- output list of links to each village page --></font>
<ul>
<?map $x in list("Les Alluets-le-Roi", "Chavenay",
                "Ecquevilly", "Noisy-Le-Roi")>
 <li>   <A>
 <?attributes attr("href",urlstr("village.rhtml","village",$x))>
   <?$x>    </A>
 </li>
</ul>
</html>
```

Fig. 5. Document prescription for a List Villages

Figure 5: Document prescription for a List of Villages

When Norfolk generates the List of Villages page, it follows the Village prescription links (e.g. `village.rhtml?village=Chavenay`) to generate the Village pages as well.

4.3 The set of prescriptions

- model 1: set of photos organised by village (village.rhtml)
- model 2: a page by (photo, context = village)
- model 3: set of photos organised by themes; each theme may have its own model.
- model 4: page by (photo, context = theme) (for the relevant photos)
- model 5: full size photo

Each model has to be designed carefully with possible variants. For example a page according to model 1 could be organised by streets if there are many photos for a village, or could display iconic photos in a unique page if there is only a small set of photos. Further links to individual photos will lead to pages generated according to model 2 or model 4, depending on the context.

It is worth to notice that model 2 and model 4 could be the same, with a parameter setting the context. It may be a quick way of prototyping the pages in a preliminary version of the CD-ROM.

4.4 Dynamic generation versus materialisation

The automatic generation of Web pages in Web-based Information systems can be either done in a dynamic mode or a compiled mode. Dynamic generation of pages arises when a page is created at the time an user tries to access it. The pages are kept virtual and never stored in a Web site. Usually the page is created by a specific program (for example a Common Gateway Interface, a CGI-script, to a database), or from a specific query language whose statements are embedded into HTML pages [16]. Another approach is to compile all the pages into a *materialised hypertext*. Although for Web sites, one can argue with [15] whether a dynamic generation is preferred to the creation of actual pages, for a CD-ROM there is no benefit in keeping the pages dynamic. The main reason is that once the information is on the CD-Rom there is no way of updating it, and therefore the generated pages will never change.

The dynamic generation is very useful though during the design and tuning of the page prescription, as well as during the process of selecting the appropriate photos, as we have explained before.

5 Comparison with other works

The main feature that distinguishes Norfolk from other approaches is that it is document-oriented. Virtual documents on the Web are usually generated through programs or scripts that sometimes can be embedded in an HTML document. This often means that the person responsible for the document (the *creator* has to rely on someone else (the *programmer* to deliver the information on the Web. Furthermore, these languages do not offer appropriate data types for the manipulation of document structures, which is crucial as virtual documents need to extract and include parts of other documents, or to render other information in a document-like manner.

The CD-ROM in [10] about architectural heritage in La-Rochelle was produced by defining XML documents (according to an Inventory DTD) which were automatically translated into HTML using XSL. XML pages were translated according to XSL templates similar to our prescriptions. The difference is that our language can query several XML documents or other data sources, while XSL templates are applied to one XML page at time. It is a one page to one page translation.

The Araneus system [1] uses a page model language to define Web pages and typed links between pages. Content of pages can be extracted from various data sources and integrated through a nested table model. Although the definition of true typed links between classes of pages is interesting we think that our tree model and page prescription language fits better with the XML/HTML models and languages.

6 Conclusion

We have shown how XML metadata can support the retrieval of photographs and the generation of advanced presentations using these photographs. The cost of creating the metadata makes it important to make them reusable for a variety of purpose and extendable to a particular application. Dublin Core can be used for photographs after some adjustments, especially on the Format and Source elements. Dublin Core does have some limitations however when it comes to describing "surrogates", or when several photographs represent the same object.

Our approach with metadata has the following advantages:

- being textual and external to the image, our metadata descriptions are easier to read and to modify.
- We can use an SGML query language to search or extract metadata (here we have used Norfolk query language).
- Our metadata scheme is extendable, as we have shown by extending the Dublic Core with qualifiers and simple schemas.

The generation of pages using virtual documents in our application has the following advantages:

- There is a considerable gain in time: instead of writing many similar pages, only one prescription has to be written.
- The pages are less prone to errors. The photograph metadata and document prescriptions, are written and checked once, and reused many times.
- There is a separation between information and delivery (ie. metadata and HTML pages), which allows to easily define several presentations of the same information
- Norfolk can support the flexible generation of presentation from heterogeneous sources, which will be a requirement for the next generation of Digital Libraries.

Future work involves the definition of models for comparison that support the generation of queries for appropriate comparable elements, and the display of these elements.

References

1. Atzeni, P., Mecca, M., Merialdo, P.: Design and Maintenance of Data-Intensive Web Sites. In *Proceedings of EDBT'98* (published as Lecture Notes in Computer Science 1377), Valencia, Spain, March 1998, pp436-450.
2. Besser, H.: *Image Databases, Database: The Magazine of Electronic Databases Review*, 1998. http://www.sims.berckeley.edu/~howard/Papers/newpapers/imagedatabase.html
3. Cox, S., *A Guide to the Dublin Core datamodel with some notations for recording Dublin Core metadata*, draft proposal, 5th May 1999 http://www.agcrc.csiro.au/projects/3018CO/metadata/dc-guide/

4. Dempsey, L., Heery, R.: *Specification for resource description methods. Part 1. A review of metadata: a survey of current resource description formats*, 1997. http://www.ukoln.ac.uk/metadata/desire/overview/
5. [DC-HOME]: The Dublin Core Metadata Initiative. Organizatioanl Website. April 1999. http:/purl.org/dc/
6. [RDF]: Resource Discovery Framework. Working Group Webpage. (April 1999) http://www.w3.org/RDF/
7. [RFC2413] Weibel, S., Kunze, J., Lagoze, C., Wolf, M.:Dublin Core Metadata for Resource Discovery.IETF Information RFC. (Septembre 1998) ftp://ftp.isi.edu/in-notes/rfc2413.txt
8. Elise: Electronic Library Image Service for Europe, Final Report, September 1995. http://severn.dmu.ac.uk/elise/
9. James, A., *Cataloguing Photographs in Archival Repositories* http://sunsite.Berkeley.EDU/Imaging/databases/fall95papers/james.html
10. La Rochelle, Patrimoine architectural et mobilier, CD-ROM, edited by the French Ministry of Culture and INRIA.
11. Miller, P., *Dublin Core Sub-elements: Summary of work prior to and during DC5, and an outline of issues still to be resolved*, Octobre 1997 http://www.mailbase.ac.uk/lists/dc-subelements/files/helsinki.html
12. Miller, E., Miller, P., Brickley D. (Eds.), *Guidance on expressing the Dublin Core within the Resource Description Framework (RDF)*, Draft Proposal, May 1999 http://www.ukoln.ac.uk/metadata/resources/dc/datamodel/WD-dc-rdf/
13. Paepcke, A., Digital Libraries: Searching Is Not Enough, D-Lib Magazine [online], May 1996. http://ukoln.bath.ac.uk/dlib/dlib/may96/stanford/05paepcke.html
14. Paradis, F., Vercoustre,A.M.: A Language for Publishing Virtual Documents on the Web. In *Proceedings of the Workshop on the Web and Databases (WebDB'98) (in conjunction with EDBT'98)*, 1998. http://www.dia.uniroma3.it/webdb98/papers.html
15. Sindoni,G.: Incremental maintenance of hypertext views. In *Proceedings of the Workshop on the Web and Databases (WebDB'98) (in conjunction with EDBT'98)*, 1998. http://www.dia.uniroma3.it/webdb98/papers.html
16. Vercoustre, A.M., Paradis, F.: A Descriptive Language for Information Object Reuse through Virtual Documents, in *4th International Conference on Object-Oriented Information Systems (OOIS'97)*, Brisbane, Australia, 10-12 November, 1997.
17. [XML] W3C: Extensible Markup Language (XML) 1.0, W3C Recommendation 10-February-1998, 1998, http://www.w3.org/TR/REC-xml
18. Weibel, S., Miller, E.: Image Description on the Internet, A Summary of the CNI/OCLC Image Metadata Workshop, Sep. 1996, D-Lib magazine, January 1997.

Appendix 1: Example of XML Metadata for the Photograph “AH3”


```
<?xml version="1.0">
<metadata lang="fr">
<DC:Title> Maison rurale à Chavenay (78), France </DC:Title>
<DC:Creator> Anne-Marie Vercoustre </DC:Creator>
<DC:Contributor> Labo Service; Art de Vivre, 78630 Orgeval,
      France  </DC:Contributor>
<DC:Date> 1990 </DC:Date>
<DC:Subject xml:lang="fr">
   maison, façade, volet, fenêtre, lucarne, toiture, ravalement
</DC:Subject>
<DC:Description xml:lang="fr">
        Charmante petite maison de village
</DC:Description>
<DC:Description SCHEME= "habitat" xml:lang="fr">
     <rural>
     <ravalement> Enduit couvrant traditionnel qui vient mourir
              sur les tableaux de fenêtres. Le soubassement est
            traité différemment pour laisser respirer la pierre.
      </ravalement>
      <fenetre> Fenêtres à 6 carreaux</fenetre>
      <toiture> Jolie toiture; Lucarne à capuche
                 qui a été raccourcie.
      </toiture>
      <volet> Volets pleins traditionnels à simple barre.
              a couleur bleue passée est pleine de charme.
      </volet>S
      <etat>  restauré,  bon </etat>
      </rural>
 </DC:Description>
<DC:Publisher> Association Sportive et Culturelle des Alluets-
       le-Roi  (LASCAR) </DC:Publisher>
<DC:Identifier>Photos/AH3</DC:Identifier>
<DC:Type> image </DC:Type>
<DC:Type> color photograph </DC:Type>
<DC:Type.quality> good </DC:Type.quality>
<DC:Format> image/mpeg </DC:Format>
<DC:Format.full> 800X537 </DC:Format.full>
<DC:Format.small> 200X134 </DC:Format.small>
<DC:Format.orientation> vertical </DC:Format.orientation>
<DC:source> AH.3 </DC:source>
<DC:Source.type> color negative </DC:Source.type>
<DC:Rights> (c)Anne:Marie Vercoustre </DC:Rights>
<DC:Relation> Habitat rural  et sa restauration, collection
```

```
        de 150 photos sur CD-ROM
</DC:Relation>
<DC:Coverage SCHEME= "address" >
       <address>
             <number>21</number>
             <street>Grand Rue</street>
             <town type=village>Chavenay</town>
             <country>France</country>
        </address>
</DC:Coverage>
<DC:Coverage>  Maule et environs, Yvelines, France
</DC:Coverage>
</metadata>
```

Appendix 2: RDF Metadata for the Photograph AH3

```
<?xml version="1.0" encoding="ISO-8859-1?>
<rdf :RDF  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
          xmlns:dc="http://purl.org/dc/elements/1.0/"
          xmlns:dcq="http://purl.org/dc/qualifiers/1.0/">
<rdf:Description rdf:about ="Photos/A3">
<dc:creator> Anne-Marie Vercoustre </dc:creator>
<dc:contributor> Labo Service; Art de Vivre, 78630 Orgeval, France
</dc:contributor>
<dc:date>
      <rdf:Description>
           <rdf:value> 1990 </rdf:value>
           <dcq:dateType> photo shot </dcq:creatorType>
           </rdf:Description>
</dc:date>
<dc:subject xml:lang="fr">
     maison, façade, volet, fenêtre, lucarne, toiture, ravalement
</dc:subject>
<dc:description>
      <rdf:Alt>
           <rdf:li xml:lang="fr">
                 Charmante petite maison de village </rdf:li>
           <rdf:li xml:lang="en"> Nice little village house</rdf:li>
</rdf:Alt>
</dc:description>
<dc:description xml:lang="fr">
      <rdf:Description>
       <rural:ravalement> Enduit couvrant traditionnel qui vient
             mourir sur les tableaux de fenêtres. Le soubassement est
                 traité différemment pour laisser respirer la pierre.
```

```
          </rural:ravalement>
          <rural:fenetre> Fenêtres à 6 carreaux</rural:fenetre>
          <rural:toiture> Jolie toiture;
           Lucarne à capuche qui a été raccourcie.
          </rural:toiture>
          <rural:volet> Volets pleins traditionnels à simple barre.
                         La couleur bleue passée est pleine de charme.
          </rural:volet>
          <rural:etat>  restauré,  bon </rural:etat>
     </ref:Description>
</dc:description>
<dc:publisher> Association Sportive et Culturelle des Alluets-le-Roi
(LASCAR) </dc:publisher>
<dc:identifier>AH3</dc:identifier>
<dc:type>
     <rdf:Bag>
          <rdf:li> image </rdf:li>
          <rdf:li> <rdf:Description>
          <rdf:value>: color photograph </rdf:value>
          <dcq:typeType> photograph </dcq:typeType>
  </rdf:Description> </rdf:li>
          <rdf:li> <rdf:Description>
          <rdf:value>: good </rdf:value>
          <dcq:typeType> quality </dcq:typeType>
          </rdf:Description> </rdf:li>
     </redf:Bag>
</dc:type>
<dc:format> image/jpeg </dc:format>
<dc:format>
      <rdf:Bag>
          <rdf:li><rdf:Description>
             <rdf:value> 800x537 </rdf:value>
             <dcq:formatType> fullsize </dcq:formatType>
          </rdf:Description> </rdf:li>
          <rdf:li> <rdf:Description>
     <rdf:value> 200x 134 </rdf:value>
             <dcq:formatType> thumbnail </dcq:formatType>
          </rdf:Description> </rdf:li>
          <rdf:li><rdf:Description>
       <rdf:value> vertical </rdf:value>
             <dcq:formatType> orientation> </dcq:formatType>
          </rdf:Description> </rdf:li>
      </rdf:Bag>
</dc:format>
<dc:source> AH.3 </dc:source>
```

```
<dc:source>
    <rdf:Description>
          <rdf:value> color negative </rdf:value>
          <dcq:sourceType> original </dcq:sourceType>
     </rdf:Description>
</dc:source>
<dc:rights> (c)Anne:Marie Vercoustre </dc:rights>
<dc:relation> Habitat rural  et sa restauration, collection de
      150 photos sur CD-ROM
</dc:relation>
<dc:coverage>
     <rdf:Description>
        <rdf:value>
<adress:number>21</number>
                <adress:street>Grand Rue</street>
<adress:town type=village>Chavenay</town>
<adress:country>France</country>
         </rdf:value>
         <dcq:coverageType> location  </dcq:coverageType>
     </rdf:Description>
</dc:Coverage>
<dc:coverage>  Maule et environs, Yvelines, France
</dc:coverage>
</rdf:Description>
</rdf:RDF>
```