Introduction to CORC Records

The CORC (Cooperative Online Resource Catalog) web site (http://www.oclc.org/oclc/research/projects/corc/) has complete information about the project. This is intended simply as a brief introduction to creating CORC records. OCLC describes CORC in this way: “CORC is a research project exploring and promoting the cooperative creation of Web resource descriptions. Because of its similarity to InterCat, we considered calling the project "InterCat2." The main difference between CORC and InterCat is that CORC integrates recent metadata initiatives with MARC. We are building a new system for CORC, so we can take a more flexible approach to record creation. In addition, CORC emphasizes exporting the records in forms usable on the Web (e.g. HTML, XML).”

CORC is designed to be viewed on a monitor with 1024 resolution. It is continually being changed with new features being added and bugs being fixed. There are more features than covered in this brief introduction, such as group export of MARC records and authority checking. CORC’s response time is quite variable, occasionally quite fast, usually longer than ideal, and occasionally extremely slow (5 minutes +).

As an example, CORC will be used to create a record for the web site in Figure 1. This web page is made primarily of graphics. This effects how well CORC can translate it.

To create a single CORC record, you choose the option “Create In Catalog”. You are then shown the screen in Figure 2, with a blank line for the URL. You can create the record in Dublin Core or MARC format. The National Trust web page will be created in Dublin Core format.

![Figure 1:](image)

![Figure 2:](image)
After you click “Create” it takes at least 30 seconds to create a record. You are put into Editing record mode.

The other possible Actions are: Display (see Figure 7), Reformat, Validate, Delete, Clone, Cancel Changes, Submit, Set Complete Status, Clear Status, Add Constant Data, Use Constant Data, Export Record DC HTML, and Export Record DC RDF.

You can also View MARC (see Figure 5), or MARC Text Area, DC HTML (see Figure 4) and DC RDF.

The pull-down menus at the left give all 15 Dublin Core Elements. The secondary pull-down menus give appropriate qualifiers for the chosen element.

The Functions in the pull-down menu at the right (Add, Control, Clear, Delete) control the existence of the field.

CORC pulls some of this information from the source code (see below), some is encoded in the web page, and some information it interprets from the web page.

The title comes from the title field in the header. This web page author used the metatags “Description” and “Keywords”. CORC places them in the first “Description” and “Subject” field, respectively.

The Date published is actually the last update date. This information is encoded in the page. CORC tries to determine the Publisher, Format, Language and Rights from the content of the page. It does not do this without error (for example, it claimed the CNN allergy page was in Danish).

The second Description usually is the first few lines of text on the page. In this case, where most of text is graphics, it pulled the description from the title field and one of the later graphics as it appears in the source code. I have no idea why it did this.

The Dewey numbers come from another experimental project at OCLC called Scorpion. The results can be irrelevant, although it has improved quite a bit during the last 6 months. This example shows that irrelevant headings that were created.

Figure 3:
You can edit either in MARC view or in Dublin Core view. There are fields in MARC that do not translate to Dublin Core. In the future, there may be fields added to the Dublin Core view that do not appear in MARC. The “Date” and “Language” fields appear in the MARC fixed fields.

Some of the translation between Dublin Core and MARC is quirky. For example, if you use the “Creator” field twice in Dublin Core (which is perfectly legitimate), CORC will change the second “Creator” to a “Contributor” (7XX). In proper Dublin Core, the order of the elements does not matter, but CORC redefines the additional “Creator” to a Secondary position.

Editing in CORC can be extremely slow, far slower than editing in our regular library catalog. It often takes ten seconds to over a minute to add a field.
After you have cleaned up the record to your satisfaction, you can view the Dublin Core (Figure 5). This can be exported (or simply copied and pasted) to place in local web pages.

When you are creating a record, CORC searches its database. If it finds a possible match, it will display a page like Figure 6. If you wish to display the page(s) that it prompts, you can click on the name of the page. (Clicking on the edit box will bring up the record in Edit mode.)

Figure 7 is an example of a record that is fully cataloged in CORC. Record quality varies tremendously. Some libraries are doing full AACR2 cataloging, other libraries are doing more limited cataloging. There are no standards for what fields are mandatory. Many libraries ignore (or remove) the subject headings because they are not helpful. In this case, the subject headings created automatically on 3/12/99 were: Law-Museums and galleries; Galleries, museums-British Isles; Nautical engineering-Sailing ships; Immaculate conception; Founders of religions. The subject headings in the existing CORC record are different, but still somewhat problematic. The headings produced by the current Scorpion system are much better: Warships; Support warships and other government ships; Power-driven warships.
CORC also does not recognize Dublin Core fields that are already in a web page. The web page in figure 8 uses Dublin Core. The source code for the Head section follows.

**Figure 8:**

```html
<HEAD>
<TITLE>U of Iowa Serials Cataloging</TITLE>
<META NAME=GENERATOR CONTENT="Claris Home Page 3.0">
<X-CLARIS-WINDOW TOP=0 BOTTOM=882 LEFT=0 RIGHT=1268>
<X-CLARIS-TAGVIEW MODE=minimal>
<meta name="keywords" content="serials cataloging Iowa University">
<meta name="distribution" content="global">
<slink rel="schema.DC" href="http://purl.org/DC/elements/1.0/">
<meta name="DC.Title" content="U of Iowa Serials Cataloging">
<meta name="DC.Title" content="University of Iowa Serials Cataloging">
<meta name="DC.Title" content="Serials Bibliographic Processing Team">
<meta name="DC.Creator" content="Christ, Ruth.">
<meta name="DC.Description" content="Home page for the Serials Bibliographic Processing Team, Central Processing Services, University of Iowa Libraries.">
<meta name="DC.Publisher" content="University of Iowa, Libraries.">
<meta name="DC.Date" content="1997-05-15">
<meta name="DC.Type" content="text">
<meta name="DC.Format" content="text/html">
<META NAME="DC.Identifier" CONTENT="http://www.lib.uiowa.edu/serials/">
<META NAME="DC.Language" CONTENT="en">
<META NAME="DC.Rights" CONTENT="Copyright (c) 1997-1999. The University of Iowa. All rights reserved.">
<META HTTP-EQUIV="Content-Type" CONTENT="text/html; charset=ISO-8859-1">
</HEAD>
```

**Figure 9:**

Note (Figure 9) that several fields (Title, Identifier, Format, Type) are duplicated. The dates are coded as if they were duplicated, although one is the date created and the other the date updated. Also note the description. The DC.Description field is not included. Instead, CORC makes its standard type of description, the first few lines of text from the page. In this case (as in most cases), the result is not very helpful. CORC did not recognize the DC fields Publisher, Creator, Language, Rights, and Subject.
Figure 10 shows a web page that uses one big frame for the entire page. CORC does not work well with frames.

The complete source code for the page reads:

```html
<html>
<head>
    <meta http-equiv="content-type" content="text/html;charset=iso-8859-1">
    <meta name="generator" content="GoLive CyberStudio 3">
    <title>Welcome to Welsh Lakes and Mountains</title>
</head>
<frameset cols="108,*" border="0" framespacing="0" frameborder="NO">
    <frame src="pages/nav.html" name="nav" noresize>
    <frame src="pages/main.html" name="main" noresize>
</frameset>

The content that you see in the page comes from the source code for the frame.

If a page such as Figure 10 uses frames, many fields will be left blank (Figure 11). CORC was not able to use any text for the description (except the Title field). CORC was not able to determine anything about the publisher or the rights, even though the page states “This site was design by Stuart Lloyd Associates, in association with Powys County Council, and was produced by Angel Multimedia.”

Prepared by: Wendy Robertson, wendy-robertson@uiowa.edu, Serials Bibliographic Processing Team, Central Processing Services, The University of Iowa Libraries


October 1999