

## STEP 1: PREPARING MSWORD DOCUMENTS FOR CODING IN OXYGEN

### BACKGROUND

To prepare and code documents you will use MSWord and oXygen.

oXygen is an xml Editor – that is, it is a software program designed to facilitate encoding of documents in xml, "validate" or check that the codes have been applied correctly, and "transform" xml documents into html by applying a style sheet (actually, two style sheets – one called a cascading style sheet (css) and a Compendium-specific (xslt) style sheet).

The glossary articles you will be encoding are currently MSWord files. oXygen requires plain text – that is, text without any formatting. The final edit Word file will have had most of the formatting already stripped out of it (except in the Works Cited and Suggested Reading lists). The main task in preparing a document for coding is to look for and replace any remaining formatting – in particular, italics and special characters with the appropriate tags or character references (entities).

Italics:

The xslt style sheet determines how titles, the author's name and affiliation, headings (including the headings "Suggested Reading" or "Works Cited" are displayed. The style sheet also looks after displaying italics in the bibl entries. What we need to pay attention to is italics in the body of an article.

There are three uses of italics that you will encounter in the body of an article and will need to tag. But first a little note about tags. You will learn specific tags as you begin coding. The important thing to know about tags for now is that they always occur in pairs – an opening and closing tag – and surround the precise text to be coded. An opening tag looks like this <tag> and a closing tag like this </tag>. Back to the three uses of italics you will encounter in the body of an article. They are:

1. Titles of books. The first mention of a book title in a paragraph is enclosed with the tag <title level="m"></title>. Any subsequent mentions of the **same** book title (full title or shortened version) are encoded with <emph></emph>.

For example:

```
<p>Naomi Klein is a Canadian journalist, media commentator, and most famously the author of <title level="m">No Logo: Taking Aim at the Brand Bullies</title> (2000), a critique of the corporate takeover of public space. Klein's youth (she was 30 when the book was published), her clear and compelling prose, and the fact that the book was released shortly before the World Trade Organization talks in Seattle all contributed to <emph>No Logo</emph> becoming a key text in the anti-globalization debate...</p>
```

2. For foreign words (not in common usage in English). These are coded with <emph></emph>.

3. Emphasized words or phrases. These are also coded with the <emph> tag.

Note: Once you have applied the appropriate tag to the italicized text, you will need to remove the italics formatting applied in Word.

#### Special Characters:

All special characters, symbols, and em-dashes need to be replaced with character entity references, or "entities" for short. Entities are a non-browser specific/non-software specific way of describing a special character or symbol. The Compendium project uses hexadecimal entities. These take the case-insensitive form **&#xnumber;**. A list of frequently used entities is attached. A full list of entities can be found at <http://htmlhelp.com/reference/html40/entities/>. To use the table found at the URL, scan the rightmost column (Rendered in your browser –Hex) for the character you want. Look across to the HEX symbol column to find the hexadecimal entity. You will need to add two zeros (00) after the &#x and before the numerals given. Remember to include the semicolon at the end. Thus the entity for the trademark <sup>TM</sup> symbol is &#x002122;

#### Now To Get Started!

1. Open the Word file.
2. Save a version of it in the appropriate "Prepped for Coding" folder.
3. Read the article to familiarize yourself with the content. Highlight all foreign characters. (There is no easy way to find these using Search and Replace in Word).
4. Search and replace all double hyphens (--) with &#x002014;
5. Search and replace all ellipsis (...) with &#x002026;
6. Search and replace all ampersands (&) with &#x0026;
7. Replace all foreign characters with appropriate entities.
8. Search for all italics (Enter ^? In Find what: Go to Format at bottom of window, select Font then choose italic). This will locate any (and every character that appears in italics). Use this to locate: (a) words and phrases that have been italicized to convey emphasis. Enclose word or phrase in <emph></emph>. And (b) titles of books mentioned in the main body of the document. Don't worry about book titles in Suggested Reading and Works Cited. Enclose titles with <title level="m"></title>. Tag the first mention of a particular book with <title level="m">. Tag all subsequent mentions of the full title or abbreviated title of the **same** book within the same paragraph with <emph></emph>. Ignore italics in Works Cited and Suggested Readings list.
9. Re-save the file.

**Please do not make changes to the text.** If you come across something that looks like a missed typo or an inconsistency in spelling (e.g., policy maker/policy-maker) or punctuation, please bring it to my attention.

## CODING NOTES

### ARTICLES WITH HEADINGS (AND SUBHEADINGS)

Headings within the article denote divisions within the body of the text. Each section or division (including the header for that section) needs to be tagged as a `<div></div>`. Headings are tagged with `<head></head>`.

Subheadings (Level 2 headings) denote divisions within a division of the main text. These divisions (including the Level 2 header) are also tagged with `<div></div>`. They represent a nested div, or a div within a div. Subheadings are tagged with `<head></head>`. The style sheet recognizes a Level 2 header as a header of a div within a div and renders it differently than a head within an un-nested div.

In order to generate the "table of contents" for each paper, which appears on the right of the screen in the Compendium, divisions of the text must be numbered. This is done by including the attribute `n=""` in the opening `<div>` tag (i.e., `<div n="1">`; `<div n="2">`). All `<divs>` must be numbered. `<divs>` within `<divs>` -- that is, subsections -- are numbered (`<div n="2.1">`; `<div n="2.2">`). Where a document has a Preface, the `<div>` is numbered as `<div n="1">`. The first `<div>` in the body of the text is numbered `<div n="2">`.

### BULLETED AND NUMBERED LISTS

Numbered lists are coded as follows:

```
<list type="ordered"><item>Universalism. Humans everywhere are seen as having similar needs and desires.</item>
```

```
    <item>Individualism. Most INGOs accept as members only individuals or associations of individuals. Their decision-making structures also this aspect, being normally based on one-person, one-vote ideas.</item>
```

```
    <item>Rational voluntaristic activity. Political action takes the form of "responsible individuals acting collectively through rational procedures" who can in turn "determine cultural rules that are just equitable, and efficient, and that no external authority is required for their legitimation."</item>
```

```
    <item>Rationalization. In engaging in rationalized and rationalizing activity, political activity seeks to achieve "rational progress."</item>
```

```
    <item>World Citizenship. Individuals vary in their capacities, resources, and industry, but all have the same basic rights and duties.</item></list>
```

The tag `<list type="ordered"></list>` goes around the entire list.

The tag `<item></item>` goes around each item in the list.

Bulleted lists are coded in the same manner but with the tag `<list type="unordered"></list>` around the entire list.

## **URLS**

URLs that appear in the body text (not the Works Cited or Suggested Reading lists) are coded as follows:

```
<xref url=http://www.ns.res.org.hn/via>www.ns.res.org.hn.via</xref>
```

Note that http:// must be given in the url that appears between the " " in the xref tag.

## Frequently Used Character Entities

Character	Character as Rendered in Browser	Hexadecimal Entities
em-dash	–	&#x002D;
horizontal ellipses or 3-dot leader	...	&#x002026;
ampersand	&	&#x0026;
Latin small letter e with grave	è	&#x00E8;
Latin small letter e with acute	é	&#x00E9;
Latin small letter c with cedilla	ç	&#x00E7;
Latin small letter a with grave	à	&#x00E0;
Latin small letter i with circumflex	î	&#x00EE;
Latin small letter n with tilde	ñ	&#x00F1;
Latin capital letter E with acute	É	&#x00C9;
Latin small letter o with circumflex	ô	&#x00F4;
Latin small letter a with umlaut	ä	&#x00E4;
Latin small letter e with circumflex	ê	&#x00EA;

Notes:

A complete list of entities can be found at:

<http://htmlhelp.com/reference/html40/latin1.html>. Most entities you will use can be found on the page for Latin-1 Entities. Locate the symbol in the column for "Rendering in Your Browser" as "Hex". Use the corresponding "Hex" reference symbol. You must add two zeros "00" before the last three characters. The last character is always a semicolon.

Always search for ampersands in your source Word file and replace with the appropriate character reference/entity **one at a time (not globally)**. Otherwise, you'll end up with entities within in your already inserted entities (for the ampersands in the entity)!

## ENTRY RULES FOR BIBLIOGRAPHIC DATABASE

Please help keep the bibliographical database in good order by following all of these guidelines. You are encouraged to print a copy of this page, for quick reference.

**Before entering a reference in the bibliographic database, you must determine, by searching the database, that the reference has not already been entered.** Failure to do so could result in multiple entries for the same reference. If your search yields an existing entry for the reference you wish to add, compare all appropriate elements (i.e., author, year of publication, title, place of publication, publisher, edition, volume and issue number, page numbers of chapters and journal articles, etc). Discrepant information may point to an error in the existing entry or the new reference. Investigate all discrepant information and if necessary edit the existing entry. As Works Cited and Suggested Reading lists are generated from the bibliographic database, it is permissible for the bibliographic database to contain separate entries for separate editions of a work and that editions may be designated by an edition number or may simply have a different publisher and place of publisher. **Works Cited lists in xml documents must contain the bibl key for the edition cited by the author.**

1. Please do not enter any **punctuation**, except exactly as and where specified in the example. Pay particular attention to final punctuation in the author field when there is a **single author**. If initials are provided, do not enter a period after the last initial. This will result in a transformed bibl entry with two periods – one following the initial and one inserted by the xslt style sheet at the end of the author field.
2. Check that you have not accidentally added **extra spaces** before or after the text in a field.
3. **Author** names ideally include a first name, initial, and last name. For **Editors** and **Translators** initials for first and second names are sufficient.
4. Text in the **Author** field should be entered as follows:
  - One author:  
Marx, Groucho L
  - More than one author:  
Marx, Groucho L. and Marx, Harpo B. (Note that there is **no comma** following the first author's name)  
Marx, Groucho L., Marx, Harpo B., and Marx, Chico (**Note the serial comma** following the second to last author's name)
5. Text in the **Editor** field should be entered as follows:
  - One editor:  
Charlie M. Chaplin
  - More than one editor:

C.M. Chaplin, G. Marx, and S. Laurel

- If the bibl is for an edited volume and no chapter is specified, enter the editor(s)' names in the Author field, followed by a period and "ed" or "eds". Do not enter a period after "ed" or "eds"; the xslt style sheet will enter the period.

C.M. Chaplin. ed

- Text in the **Translator** field should be entered according to the same rules as text in the Editor field.

Charlie M. Chaplin

6. **Titles of Books, Book sections (chapters) or Journal articles** should be entered as follows:

- Title with no subtitles:  
Truth flies like an arrow
- One Title and one subtitle:  
Truth flies like an arrow: Fruit flies like a banana
- One Title and two subtitles:  
Groaning with Groucho Marx: Truth flies like an arrow: Fruit flies like a banana

There are no spaces before a colon. Note that all titles are set in **sentence case** – that is, only the first word (and the first word following a colon) are capitalized. **Proper nouns** are an exception to this rule and should be capitalized.

**Titles of Journals** should be entered in title case (e.g., Canadian Journal of Sociology).

Do not enter a period after a title. The xslt style sheet will enter the proper punctuation (in this case, no punctuation).

If an **apostrophe** appears in the title (or elsewhere in the bibl), you must enter as "\'", placing a slash before the apostrophe.

7. Text in the **Year of Publication** field should be entered as follows:  
YYYY.

For reprints enter YOP as follows 1896/2001 where the first date is original YOP and second date is the reprint date.

8. Text in the **Date Month/Day** field should be entered as follows (examples with and without day numbers):

21 March

April

9. Text in the **Pages** field should be entered as a number, or as numbers with a hyphen between them for a page range. Do not enter "p." or "pp." before the page number: 55-67

Do not enter a period after the page number or page range. The xslt style sheet will enter the proper punctuation.

Page ranges should be indicated as follows: 1-17; 43-9; 43-54; 110-7; 118-54; 1372-9; 1385-94; 1399-1405.

10. Text in the **Volume Number** and **Issue Number** fields should be entered as a number with no punctuation:

103

If the issue or volume has a number, there is no need to enter the associated season (if listed). However, if there is no issue number or volume number, and there is another designation such as an associated season, you can enter it as follows:

Spring

11. Text in the **Place of Publication** field should be entered as follows: For well-known metropolitan cities (Chicago, New York, Toronto, San Francisco, London, Paris), it is not necessary to list the state or province. For other places of publication within North America, list the city and the two-letter abbreviation for the province or state (see *Chicago Manual of Style*, 15<sup>th</sup> ed. Section 15.29). For less-known places of publication outside North America, list the city and the country.

Boulder, CO

New York

Ostfildern, Germany

12. In the **Date Accessed** field enter the date as follows:

7 June 2005

13. For **Working Papers**, enter the information in the **Author**, **Year of Publication** and **Title of Paper** field as you would for other published works. Use the **Publication Statement** field to enter a statement that gives any document identification information along with the "publisher." See the example below:

Author: Scholte, Jan Aart

YOP: 2002

Title of paper: What is globalization? The definitional issue -- again

Publication Statement: CSGR Working Paper No. 109/02. Centre for the Study of Globalisation and Regionalisation, University of Warwick, UK

Note the **period** following the document number. Do **not** enter a period at the end of the statement. The xslt will insert it.

14. Unpublished Papers are papers presented at conferences or meeting that are not/have not yet been published. Papers published as part of a set of conference proceedings are treated as book sections/chapters. For Unpublished Papers, enter the information in the **Author**, and **Title of Paper** field as you would for other published works. Enter the year in which the conference was held in the **Year of Conference** field. Use the **Name and Place of Conference** field to enter a statement that gives the name, location and full dates of the conference. See the example below:

Author: Via Campesina

Year of Conference: 2000

Title of Paper: Biodiversity, biosafety and genetic resources

**Name and Place of Conference:** Paper presented at a meeting in Bangalore, India, 15-18 October 2000

Do not enter a period at the end of the statement. The xslt will insert it.

## **PROCEDURE FOR CODING GLOSSARY ARTICLES WITHOUT BIBLS (AND RESEARCH SUMMARIES)**

1. Open "prepped for coding" version of the Word file.
2. Open glossary template in Oxygen.
3. Save as: Provide document name, but place in Coded folder.  
File name conversions:  
For Research Summaries, Position Papers, and Research Articles the convention is RS\_Cook\_DevelopingCommunity. The first two letters indicate the document type (RS, PP, RA); followed by underscore then Author(s)' surname; followed by underscore and the title keyword. The final draft Word file should already follow the convention. You can just reuse it. Glossary Articles follow the convention: GL\_CO\_Citizenship where GL indicates Glossary; second two letters indicate the type of gloss (CO = concept; PL = place; EV = event; PR = person; and OR = organization). For the last word please use whatever I've named the Word file.
4. Complete TEI header information:
  - Copy and paste the title, author, and author affiliation from Word
  - Enter synonym(s)/acronym; If none, delete <title type="synonym">Synonym or acronym</title>. If more than one, repeat tag for each.
  - Enter encoder name, date value and date in display format
  - Enter topics.
5. Assign text id. Record Text ID on Glossary ID sheets in Tracking Tools folder.
6. Copy body text of article from Word, highlight text between <div></div> and paste Word text between <div> tags.
7. Tag paragraphs. Easy way to do this is to highlight paragraph. Right click. Select "xml refactoring" from list of options, then select "surround with tag", then "p". Tip: Ctrl/ will allow you to tag selected text with the last tag used.
8. Since the article does not contain either Suggested Readings or Works Cited, delete <back>

```
      <!-- if glossary entries do not contain works cited or suggested readings, delete
this section -- >
      <div type="bibliography">
        <listBibl type="wc">
          <!-- Works cited -->
          <bibl id="OR.0111.unesco1945" key="unesco.1945"/>
        </listBibl>
        <listBibl type="sr">
          <!-- suggested readings -- >
          <bibl id="OR.0111.unesco1988" key="unesco1988"/>
        </listBibl>
      </div>
    </back>
```

Note that the template for Research Summaries does not contain <back></back>.
9. Save file.

10. Validate file.

11. Fix any errors. Error messages will help you locate and identify errors. Tip: Validate after correcting each error. Sometimes correcting one thing will simultaneously clear out other error messages.
12. Validate file a final time.
13. Save file.

## **PROCEDURE FOR CODING DOCUMENTS WITH BIBLS**

### **NOTES**

1. A bibl is Compendium lingo for an entry in the Bibliographic Database, Works Cited (WC) list, or Suggested Reading (SR) list.
2. Only Glossary Articles will have a Suggested Reading list.
3. Each bibl item (whether a Suggested Reading or a Work Cited) in an xml document contains two pieces of information: an id and a key. An important exception to this is a website listed as a Suggested Reading. These do not have a key. The key points to the appropriate entry in the bibliographic database that will be automatically plunked into the xml document once the file is uploaded. Websites (as such) are not included in the bibliographic database, so this info needs to be entered "by hand" for this type of SR. Instructions for entering websites as SR are at the end of the "Steps" section. The id has two functions: Its first function applies to Works Cited lists only. That function is to mediate the link between the in-line citation and the key that points to the bibliographic database. The id is the "target" of the `<ref target=""></ref>` tag that surrounds the in-line citation. The second function of the id is to allow the xslt to sort the bibl items in the correct order in both the Works Cited and Suggested Reading list. This means that the bibls do not have to be listed in order in the xml document by the encoder (but I find it makes sense to enter them in the order that they appear in the edited Word version of the file). It also means that ids have to be carefully defined with sorting in mind.

### **STEPS**

1. Follow the steps outlined for preparing documents to be coded. It is not necessary to tag italicized text in the bibls with `<emph>`. The xslt will take care of rendering what needs to be italicized. It will be helpful for you to insert entities for special characters though.
2. Search the bibliographic database to determine which bibls from the WC and/or SR lists need to be added and which are already there. Record the bibl keys of those that are already in the database. Enter the bibls that are not already in the database following the "Bibliographic Entry Rules." Record the bibl keys for the newly entered items. Keep track of any websites that appear in the SR lists of Glossary Articles. They will need to be included in the xml document.
3. Go to the Excel "Bibliographic Database Keys in Use" file and record the bibl keys used by the document. If a document uses a key for an item that was already in the database, but to-date the bibl hasn't been used by a Compendium document, you will need to insert a row in the list of bibl keys. Keys for newly added bibls will go at the end of the list. If the document uses a key for an item that was already in the database and is used by one or more other documents you will need to add the name of current doc in the appropriate column.
4. Open document template in Oxygen.
4. Save as: Provide document name

Use same file naming convention outlined in Step 3 of "Procedure for Coding Glossary Articles without Bibls (and Research Summaries), and place in Coded Files folder

5. Complete TEI header information:
  - For Glossary Articles indicate the type of gloss (i.e., concept, person, place, organization, event) in the <TEI.2 type> tag at the start of the header (e.g., <TEI.2 type="glossary.concept">
  - Title, author, author affiliation
  - For Glossary Articles, enter synonym(s)/acronym. The coder will need to determine these. If more than one, repeat tag for each.
  - Enter encoder name, date value and date in display format.
  - Key words. Research Articles and Position Papers employ keywords. These are entered as follows:

```
<term type="subject">poverty</term>
<term type="subject">modernization</term>
<term type="subject">market fundamentalism</term>
```
  - Topics. Like Research Summaries, Glossary Articles, Research Articles and Position Papers are indexed to one or two of nine topics. These are entered as follows:

```
<term type="topic">Trade and Finance</term>
<term type="topic">Property Rights</term>
```
6. Assign Text id. For Position Papers, Research Articles, and Research Summaries the text id will be of the format: AuthorSurname.TitleKeyword (e.g., <text id="Sneyd.ClinicalEconomics">). For Glossary Articles the text id will be of the format: <text id="CO.0049">. The convention for naming glossary ids is OR. for organization; PR. for persons; PL. for places; EV. for events; and CO. for concepts. The number value of id is assigned based simply on the order in which the type of gloss is encoded. Go to appropriate MSWord Glossary IDs tracking sheet to determine id of next article. Every article must have a unique id. The id is used to create bibl ids (e.g. OR.0111.unesco1945; where unesco = (corporate)author and 1945 = YOP or OR.0111.JohnsonN2001 where JohnsonN = author surname and initial(s) and 2001 = YOP.
7. Copy and paste main body text of article. Apply <div>, <p>, <list>, etc tags. Remember divs need to be numbered. When a document has a preface, the preface goes in <front></front> and is <div n="1">.
8. Does the article contains Works Cited?
  - No. Delete <listBibl type="wc">

```
<!-- Works cited -->
<bibl id="OR.0111.unesco1945" key="unesco.1945"/>
```

</listBibl> from template.  
Go to Step 10
  - Yes.
    - a. Count the number of bibls in the Works Cited list in the Word version. Go to Oxygen. Click in the <bibl id="" tag and insert the text id (eg., OR.0011 or Sneyd.ClinicalEconomics) as the bibl id. Copy the entire <bibl id=""

key=""></bibl>tag . Insert cursor after tag and paste (x number of bibls minus one) so that you now have the same number of <bibl id=""> tags as you have number of bibl items. Place each <bibl id=""> on a separate line. Check that you've created the right number of bibl ids.

b. Create the rest of each <bibl id=""> following the convention used in the example:

(e.g. OR.0111.unesco1945; where unesco = (corporate)author and 1945 = YOP or OR.0111.JohnsonN2001 where JohnsonN = author surname and initial(s) and 2001 = YOP). Note capitalization and that there is no punctuation and no spaces). If there is more than one author, convention is first author's surname and initials, second author surname and initials, third author surname and initials, (if more than three authors then et al after third), YOP

c. Using the list of key numbers you recorded for the document, insert the key with the corresponding id in the <bibl> tag (e.g., <bibl id="CO.0049.BalibarE2004" key="4161"/> and <bibl id="Sneyd.ClinicalEconomics.BatesR2005" key="4630"/>)

d. Go to the top of the body text (in Oxygen) and click into the start of the document. Use the Find feature to sequentially locate in-line citations. Search for opening parentheses "(" . All citations are enclosed in ( ). Select text within and including parentheses. Tag with <ref target=""></ref>. Where parentheses contain more than one citation, surround each citation with a ref target tag -- exclude the parentheses and semicolon separating citations). A ref target is a pointer tag. The tag points to the bibl id in the Works Cited list, creating a clickable link in the online version of the document. Name the ref target id=. The ref target id for <bibl id="Sneyd.ClinicalEconomics.BatesR2005" key="4630"/> will be <ref target="Sneyd.ClinicalEconomics.BatesR2005">. Note that the same work may be cited more than once in a document, but all mentions of that work -- including (e.g., Ibid., 3) -- will have the same ref target=. Each in line citation needs a ref target tag but only each **unique** ref target name needs a corresponding bibl id. You will need to click into the document at the end of the closing </ref> tag and then click the Find tool again to look for the next open parenthesis after filling in the bibl id. Note that working sequentially through the document in this way, searching on "(" allows you to find each inline citation and create/verify an entry in the works cited list.

9. Does the article contain Suggested Readings?

Note that Research Articles, Position Papers and Research Summaries will not contain SR.

No. Delete <listBibl type="sr">

<!-- Suggested readings -->

<bibl id="OR.0111.unesco1988" key="unesco1988"/>

</listBibl>

Yes. Follow Steps 8a-8d above.

10. Save file.
11. Validate file.
12. Fix any errors. Error messages will help you locate and identify errors. Tip: Validate after correcting each error. Sometimes correcting one thing will simultaneously clear out other error messages.
13. Validate file a final time.
14. Save file.
15. If the document being coded is a Research Article, Research Summary or Position Paper, you can tag it. **DO NOT TAG GLOSSARY ARTICLES OR POST THEM.** Adding a new Glossary Article to the Compendium requires that the rest of the Compendium content be tagged to it. For this reason we accumulate Glossary Articles and post them in batches using an updated taglist to tag them. Then we go back and "hand tag" mentions of the new glossary terms in the existing Compendium content.

### **ENTERING WEBSITES AS SUGGESTED READINGS**

Websites listed as SR have a bibl id so that they will sort along with other SR bibls. Instead of a key=" ", the pertinent information (information to be displayed) is entered by hand as follows:

```
<bibl id="CO.0004.SlowFoodUSA"><name> Slow Food website</name><emph>Slow  
Food International Manifesto</emph><xref  
url=http://www.slowfoodusa.org/about/manifesto.html>www.slowfoodusa.org/about/ma  
nifesto.html</xref><note>(accessed 11 March 2005).</note></bibl>
```

Note that the website "owner" is used in the bibl id in the same way that a personal author's name is. This particular website SR directs the reader to a specific webpage. The title of the webpage is given after the name of the website and is enclosed with the <emph> tag. Most website SR just have the name of the website followed by the URL and an access statement.

## PROOFING TERM TAGS

The tagger is a helpful but not very sophisticated tool. It will tag all mentions of a string of text within an article. It is the `<term target>` tag with the attribute `n="1"` that will be rendered as a link. The tag list defines separate patterns for related terms (alternative capitalization, plural/singular): e.g., genetically modified organism, genetically modified organisms, Genetically modified organisms, Brazilian Landless Worker's Movement and Movimento dos Trabalhadores Rurais Sem Terra.

The `n=""` attribute was added to the `<term target>` tag to allow us to recognize the first mention of a term and to render it as a link. We subsequently discovered that, for reasons the programmer wasn't able to determine and fix, in very large files the xslt was "crapping out" and randomly not displaying the content of certain term tags. To fix the problem on the coding end (over the programming end), we need to remove the `<term target>` tags where `n="2"` or more so that a file only contains `<term target>` tags where `n="1"`. It makes the attribute redundant but this is what we're stuck with. Removing the "extra" `<term target>` tags is one step in proofing the tags. The other is to look for and remove multiple tags for variants of a term. For example, the tagger might find and tag both "Brazilian Landless Workers' Movement" and "Movimento dos Trabalhadores Rurais Sem Terra" in a document. In this case, preserve only the `<term target>` tag surrounding the first mention of either term and delete all subsequent `<term target>` tags. The third part of proofing the tagging is to look for inappropriate applications of the `<term target>` tag. This happens most often with glossary terms that have a broad/general meaning beyond the meaning used in the context of globalization and autonomy. These are glossary terms like "place," "community," or "identity." You may need, in some instances, to remove a `<term target>` tag. The fourth part of proofing the tagging is to look for words or phrases that should be rendered as a link to a glossary article, but the tagger failed to tag. For example, we have a glossary article on the WTO Ministerial meeting in Cancun (we also have one on the WTO Ministerial meeting in Seattle). It is difficult to define a set of text strings that will catch all the ways that authors use to talk about the event. You may need to insert a `<term target>` tag "by hand". See the lists of Glossary IDs. Remember to include `n="1"`. The third and fourth parts of proofing will be a bit difficult for you until you gain familiarity with the Compendium glossary content. No worries. Between the two of us, we should be able to catch most things.

Re-save file.

## TAGGING

This step in the coding process establishes the links between the glossary articles and mentions of the defined term. It makes use of a tool developed by the McMaster TAPoR project. The tool works by looking for defined strings of text and applying the tag `<term target=" " n=" ">` (e.g., `<term target="OR.0049" n="1">`). OR.0049 in the example denotes the ID for the glossary article that the link points to. The "1" in `n="1"` denotes the first occurrence of the text string found by the Tagger. The tag list (see Tag Source below) defines the text strings that the tagger should search for and apply the `<term target>` tag to.

Go to: <http://taporware.mcmaster.ca/~taporware/otherTools/tagger.shtml>.

Enter Text Source: Browse local directory for file to tag.

Enter Tag Source: Browse local directory and select tag list file.

For Glossary Articles only you will need to enter a term NOT to apply tags to. For example, if you are tagging the glossary article for Democracy you will want to tell the tagger not to find mentions of the term "democracy" within this file and you would enter "democracy" where it says "Submitted Term (No Tagging Required)." The term entered here is case insensitive, so you could enter either "democracy" or "Democracy."

Click "Submit." Results will appear on your screen.

Use Edit/Find in this page and search for "`<term target`" in your browser to "pre-proof" the results – that is, to see if the tagger did its job and applied tags.

Edit/Select all to highlight results. Copy.

Go to oXygen and open the Dummy Tagger File in the XML Coding Tools folder.

Save the file in the appropriate "Tagged" folder with the same name as the coded file (e.g., RS\_EatonPorter\_Accounting).

In oXygen, click Edit/Select All, then Paste. This will paste the results from the tagging operation into your new xml file. Verify that the results of the tagging operation (displayed in your browser) were indeed pasted into oXygen, overwriting the content of dummy tagger file.

Re-save the file.