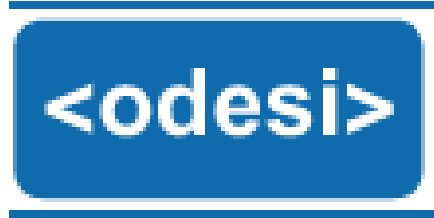# Best Practices Document
Based on DDI 2.x

Version 2

odesi.ca

August 2008

Jane Fry (Carleton University)
A. Michelle Edwards, PhD (University of Guelph)
Alexandra Cooper (Queen's University)

ODESI Best Practices Document: Version 2
Based on DDI 2.x

Table of Contents

1.1   Introduction

Ontario Data Documentation, Extraction Service and Infrastructure Initiative (ODESI), is a jointly funded project between OntarioBuys and the Ontario Council of University Libraries (OCUL). It provides university students and researchers with access to datasets in Nesstar, a web-based data extraction system. These datasets are delivered through Scholars Portal.

The goal of this document is to share our DDI knowledge with other academic institutions. It is also to include the best practices for DDI so there will be consistency in marking up the surveys.

We would like to emphasize that this is a DDI document. For this reason, any Nesstar-specific notes are included in Appendix B. If you are using Nesstar, we encourage you to read this appendix before you begin marking up your survey. We would like to emphasize here that DDI and Nesstar are two entirely different entities. Nesstar is dependent upon DDI, however, DDI is not dependent on Nesstar. Nesstar is merely a tool used to interpret DDI. It should be noted that other programs are available that create partial or complete DDI compliant codebooks, eg. SDA.

Please note that depending on which version of DDI you are using (1.x vs 2.x) some of the tags may have different numbers, but the content remains the same.

1.2   The Main Document

1.2a   The Set-up of Sections 1 to 5

The main part of this document, which follows the introduction, is composed of 5 sections.
- Section 1 – Document Description
  - describes the electronic document that the individual creates by using the DDI – the resulting XML file. The description will deal with the created DDI file and not the contents of the data or any description about the study.

- Section 2 – Study Description
  - describes the study or survey that was conducted. Information regarding purpose, how, why, where, how long, funding of the study. This section does NOT deal with the data file, that is, it only describes the study.

- Section 3 – Data Files Description
  - describes the data files that were created by the study. Description will include items such as the number of variables, the number of cases, and structure of the datafile.

- Section 4 – Variable Description
  - describes the variables included in the datafile. Information such as the variable label, the values and value labels of the variable, the literal question, who was asked the question, and any interviewer instructions are typical in this section.

- Section 5 – Other Documentation
  - allows the user to include any associated documentation. This will often include the questionnaire used to collect the data, a user guide, or a codebook.

1.2b   The Set-up of Each Section

Each section starts with a brief definition and a list of the Document Type Definition (DTD) numbers and the corresponding tags. This is followed by an in-depth description of each tag and working examples, when necessary. Then at the end of each section there is an example of a complete DDI compliant codebook in XML for that particular section.

1.2c   The Set-up of each Tag

The first line in each tag has:
- the number of the section;
- what the tag looks like in DDI; and
- the title of the tag.

> *Example:*
> > **1.1**   <citation>   Bibliographic Citation

This is followed by 3 bullets.
- Mandatory <u>or</u> Optional
  - if this tag is mandatory, use it if you have the information
  - if this tag is optional, use it where applicable
- Repeatable <u>or</u> Not Repeatable
  - if it is repeatable, this tag can be used as many times as necessary
  - if it is not repeatable, this tag is used only once
- Attributes (ID, xml:lang, source, …)
  - provide additional information about the tag
  - often provide information that is not part of the data

The description of the tag is next. This varies in length depending on the tag itself. Any general notes or formatting notes follow the description. And then the last part of the tag contains an example(s), where applicable.

Please note that the tags are filled in according to information available. This means that all the tags listed will not necessarily be used. As well, there may be tags not listed here that you deem necessary for your document.

Also note that some of the examples include the tag for the title of the survey. This is done to make the example clearer.

1.3   Extra Hints for Tags

1.3a   Formatting Conventions

Anywhere there is a formatting note, the icon below will be shown so your eye will easily catch the notes. It is important to follow these formatting conventions as these surveys will be shared with others and consistency is of the utmost importance.

- Date:
  > yyyy
  > *Example:* 2007

  > yyyy-mm-dd
  > *Example:* 2005-01-25

- Personal Names:
  > Last name, First name
  > *Example:* Edwards, Michelle

- Agency/Institution and Place of Work/Department:
  Institution. Place of work (yes, that is a period between the two)
  Example: University of Guelph. Data Resource Centre
  Example: Statistics Canada. Data Liberation Initiative

- Personal Names with Institutions/Agencies:
  Last Name, First Name; Institution. Workplace
  Example: Edwards, Michelle; University of Guelph. Data Resource Centre

1.3b   Abbreviations and Acronyms

Abbreviations and acronyms should **not** be used for anything, that is, province, country, university name, place of work, people's names, …

The only exception is if the tag is specifically for abbreviations and/or acronyms.

We have to remember that the tags are independent of each other so even if a certain institution has been listed with its full name in another section of the document, a user may only be interested in the tag at the end of the document and may not know the abbreviation or acronym. As well, the user may not be familiar with the abbreviations and acronyms that we use. We have to remember who our end user is.

1.3c   Note re: Notes and Comments Tags

The user will see that some sections include a "Notes" tag at the end. However, this tag can be included at the end of almost any section, if necessary.

This tag would include anything that is noteworthy about the previous section that has not already been mentioned. If the use of this tag is necessary, check the DDI tag document that is sourced at the beginning of each section for the correct number to be used.

See Appendix F for more information on these tags.

1.3d   Obtaining Information for Tags

When you are inputting the metadata, you will need to go searching for some of the information. For Statistics Canada data, check out Appendix C. For other data, eg. public opinion polls, you will need to look in the user guide, the background document, the methodology, …

Some of the tags you may need to do this for are:
2.2.3.2 Date of Collection;
2.3.1.1 Time Method;
3.1.3   File Structure;
3.1.6   Data Format; and
4.3.1   Location

1.4   A Note on DDI Compliant Codebooks

True DDI format creates one DDI compliant codebook for multiple data files of one survey. For example, the 2006 Canadian Tobacco Use Monitoring Survey (CTUMS) has 6 data files associated with it. True DDI format would create one DDI compliant codebook that contains Section 1, 2, and 5 describing the survey, as well as Sections 3 and 4 that contain information tags for the 6 data files. However, commercial products currently available are limited in this regard. Therefore, one DDI compliant codebook is created for each individual data file associated with the survey. For example, with the 2006 CTUMS, there will be 6 DDI compliant codebooks created.

1.5   Questions and/or Comments

If you have any questions and/or comments about this document, please send them to the <odesi> Developers (A. Michelle Edwards, PhD, Jane Fry) at odesi@uoguelph.ca

1.6   Acknowledgements

The format of this document is based on the DDI Version 2.1 XML Schema Tag Library, found at http://www.ddialliance.org/DDI/dtd/version2-1-all.html

The primary authors of the <odesi> Best Practices Document are A. Michelle Edwards, PhD, University of Guelph, and Jane Fry, Carleton University. They would also like to thank Alexandra Cooper from Queen's University, Carol Perry, University of Guelph, and Jeff Moon, Queen's University for their invaluable questions and comments.

1.7   Project Funding

<odesi> is a jointly funded project between OntarioBuys (a program of the BPS Supply Chain Secretariat, Ontario Ministry of Finance), http://www.fin.gov.on.ca/english/ontariobuys/ and the Ontario Council of University Libraries (OCUL), http://www.ocul.on.ca/ . As well, many universities have provided 'in kind' help.

SECTION 1.0   DOCUMENT DESCRIPTION   <docDscr>

The document description section of this file is intended for information regarding the marked up document itself. In other words, this section describes all the information about the metadata, or "codebook" for this file.

The Document Description, or codebook header, consists of bibliographic information describing the marked-up codebook, or document instance, itself; it contains the basic elements necessary to uniquely identify the marked-up codebook. Note that the Document Description also contains a citation section for the source documentation, that is, the hardcopy or electronic codebook that served as the source for the marked-up codebook.

Tags and corresponding DTD numbers are from DDI Version 2.0 codebook.
Source: DDI Codebook
http://www.icpsr.umich.edu/DDI/dtd/version2-1-all.html
http://www.icpsr.umich.edu/DDI/dtd/version2-1-tree.html

DTD Numbers	Tags
1.0	<docDscr>
1.1	<citation>
1.1.1	<titlStmt>
1.1.1.1	<titl>
1.1.1.2	<subTitl>
1.1.1.3	<altTitl>
1.1.1.4	<parTitl>
1.1.1.5	<IDNo>
1.1.2	<rspStmt>
1.1.2.1	<AuthEnty>
1.1.3	<prodStmt>
1.1.3.1	<producer>
1.1.3.2	<copyright>
1.1.3.3	<prodDate>
1.1.3.4	<prodPlac>
1.1.3.5	<software>
1.1.3.6	<fundAg>
1.1.3.7	<grantID>
1.1.4	<distStmt>
1.1.4.1	<distrbtr>
1.1.4.2	<contact>
1.1.5	<serStmt>
1.1.5.1	<serName>
1.1.5.2	<serInfo>
1.1.6	<verStmt>
1.1.6.1	<version>
1.1.6.2	<verResp>
1.1.6.3	<notes>
1.1.7	<biblCit>
1.1.8	<holdings>
1.3	<docStatus>

Description of Tags and Working Examples

1.0   < docDscr >   Document Description

- Optional
- Repeatable
- Attributes: <u>ID, xml:lang, source</u>

Description: The Document Description consists of bibliographic information describing the DDI-compliant document itself as a whole. This Document Description can be considered the wrapper or header whose elements uniquely describe the full contents of the compliant DDI file. Since the Document Description section is used to identify the DDI-compliant file within an electronic resource discovery environment, this section should be as complete as possible. The author in the Document Description should be the individual(s) or organization(s) directly responsible for the intellectual content of the DDI version, as distinct from the person(s) or organization(s) responsible for the intellectual content of the earlier paper or electronic edition from which the DDI edition may have been derived. The producer in the Document Description should be the agency or person that prepared the marked-up document. Note that the Document Description section contains a Documentation Source subsection consisting of information about the source of the DDI-compliant file-- that is, the hardcopy or electronic codebook that served as the source for the marked-up codebook. These sections allow the creator of the DDI file to produce version, responsibility, and other descriptions relating to both the creation of that DDI file as a separate and reformatted version of source materials (either print or electronic) and the original source materials themselves.

1.1   <citation>   Bibliographic Citation

- Optional
- Not Repeatable
- Attributes: <u>ID, xml:lang, source</u>, MARCURI

Description: Bibliographic information describing the electronic codebook, including title information, statement of responsibility, production and distribution information, series and version information.

A MARCURI attribute is provided to link to the MARC record for the citation

1.1.1   <titlStmt>   Title statement

- Mandatory
- Not Repeatable
- Attributes: <u>ID, xml:lang, source</u>

Description: Title statement for the work at the appropriate level: marked-up document; marked-up document source; study; study description, other materials; other materials for study.

     Contains the following sub-elements:

1.1.1.1   <titl>   Title

- Mandatory
- Not Repeatable
- Attributes: <u>ID, xml:lang, source</u>

Description: Title of marked up document. The marked-up codebook title will in most cases be identical to the title for the data collection (2.1.1). A full title should indicate the geographic scope of the data collection as well as the time period covered. Equivalent to Dublin Core Title.

Formatting Note 1:
> If the country name is in the title of the survey, it does not need to be put in square brackets at the end of the title statement.

Formatting Note 2:
> No abbreviations or short forms should be used in the title, eg. name of province, country.

Example 1:
> <titl>**Sun Exposure Survey, 1996 [Canada]**</titl>

Example 2:
> <titl>**Household Internet Use, 2001 [Canada]**</titl>

Example 3:
> <titl>**General Social Survey, 2005 [Canada]: Cycle 19, Time Use, Main File**</titl>

Example 4:
> <titl>**Canadian Tobacco Use Monitoring Survey, 2004: Cycle 1, Household File**</titl>

Example 5:
> <titl>**Canadian Community Health Survey, 2005: Cycle 3.1, Health Sample File**</titl>

Example 6:
> <titl>**Canadian Gallup Poll, May 1949, #186**</titl>

Example 7:
> <titl>**Domestic Violence Experience in Omaha, Nebraska, 1986-1987 [United States]**</titl>

1.1.1.2   <subTitl>   Subtitle

- Optional
- Repeatable
- Attributes: <u>ID, xml:lang, source</u>

Description: A subtitle is a secondary title used to amplify or state certain limitations of the main title.

Note:
> The subtitle includes everything after the colon in the title.

Example 1:
<titl>**General Social Survey, 2005 [Canada]: Cycle 19, Time Use, Main File** </titl>
<subTitl> **Cycle 19, Time Use, Main File**</subTitl>

Example 2:
<titl>**Canadian Tobacco Use Monitoring Survey, 2004: Cycle 1, Household File** </titl>
<subTitl> **Cycle 1, Household File**</subTitl>

Example 3:
<titl>**Canadian Community Health Survey, 2005: Cycle 3.1, Sub-sample 1: Health Utility Index (HUI), Fruit and Vegetable Consumption and Labour Force (long form)** </titl>
<subTitl>**Cycle 3.1, Sub-sample 1: Health Utility Index (HUI), Fruit and Vegetable Consumption and Labour Force (long form)**</subTitl>

1.1.1.3   <altTitl>   Alternate Title

- Optional
- Repeatable
- Attributes: ID, xml:lang, source

Description: The alternative title may be the title by which a data collection is commonly referred to or it may be an abbreviation for the title.

<odesi> format

Formatting Note:
      If there is a subtitle after the title, it should be included in the alternate title in full.

Example 1:
<altTitl>**SES 1996**</altTitl>

Example 2:
<altTitl>**HIUS 2001**</altTitl>

Example 3:
<altTitl>**GSS19 2005: Cycle 19, Time Use, Main File**</altTitl>

Example 4:
<altTitl>**CTUMS 2005: Cycle 1, Household File**</altTitl>

Example 5:
<altTitl>**CCHS 2005: Cycle 3.1, Health Sample File**</altTitl>

1.1.1.4   <parTitl>   Parallel Title

- Optional
- Repeatable
- Attributes: ID, xml:lang, source

Description: Title translated into another language.

Example:
  <titl>**Survey of Household Spending, 2002 [Canada]**</titl>
  <parTitl>**Enquête sur les dépenses des ménages, 2002 [Canada]**</parTitl>

1.1.1.5   <IDNo>   Identification Number

- ▪ Optional
- ▪ Repeatable
- ▪ Attributes: <u>ID, xml:lang, source</u>, agency, level

Description: Unique string or number (producer's or archive's number) for the marked-up document. An "agency" attribute is supplied. Equivalent to Dublin Core Identifier.

Note 1:
  This ID number is the same for the document description and the study description, ie. 1.1.1.5 is the same as 2.1.1.5.

Note 2:
  For Statistics Canada surveys, the catalogue number refers to the microdata file.

<odesi>
format

Formatting Note 1:
- • Language: E = English, F = French
- • Year: yyyy or yyyy-mm-dd
- • Lowercase: should be used for everything except the catalogue number and the language abbreviation
- • Surveys with cycle numbers and sub-numbers: use a dash between the numbers and not a period, eg. for cycle 2.1 the format would be c2-1
- • ICPSR surveys: use their ID# and add the short form if there is a subset, eg. icpsr9721im

<odesi>
format

Formatting Note 2:
  This is the format to be used:
    acronym_CatalogueNumber_language_year_subset

Example 1:
  <titl>**Sun Exposure Survey, 1996 [Canada]**</titl>
  <IDNo>**ses_82M0019_E_1996**</IDNo>

Example 2:
  <titl>**General Social Survey, 2005 [Canada]: Cycle 19, Time Use, Main File**</titl>
  <IDNo>**gss_12M0019_E_2005_c19_main-file**</IDNo>

Example 3:
  <titl>**Canadian Tobacco Use Monitoring Survey, 2004: Cycle 1, Household File**</titl>
  <IDNo>**ctums_82M0020_E_2004_c1_household-file**</IDNo>

Example 4:
  <titl>**Canadian Tobacco Use Monitoring Survey, 2004: Cycle 1, Household File**</titl>
  <IDNo>**cchs_82M0013_E_2005_c3-1_health-sample**</IDNo>

Example 5:
  <titl>**Canadian Gallup Poll, May 1949, #186**</titl>
  <IDNo>**cipo_186_E_1949-05**</IDNo>

Example 6:
<titl>**Descriptors and Measurements of the Height of Runaway Slaves and Indentured Servants in the United States: Runaway Indentured Servant Height Measurements, 1700-1850**</titl>
<IDNo>**9721im**</IDNo>

1.1.2   <rspStmt>   Responsibility statement

- Optional
- Not Repeatable
- Attributes: <u>ID, xml:lang, source</u>

Description: Person/persons responsible for marking up the document.

Contains the following elements;

1.1.2.1   <AuthEnty>   Author Entity

- Optional
- Repeatable
- Attributes: <u>ID, xml:lang, source</u>, affiliation

Description: The person, corporate body, or agency responsible for the work's substantive and intellectual content. Repeat the element for each author, and use "affiliation" attribute if available. Invert first and last name and use commas. Author of data collection (2.1.2.1) maps to Dublin Core Creator element. Inclusion of this element in codebook is recommended.

The "author" in the Document Description should be the individual(s) or organization(s) directly responsible for the intellectual content of the DDI version, as distinct from the person(s) or organization(s) responsible for the intellectual content of the earlier paper or electronic edition from which the DDI edition may have been derived.

Note:
If a student is working on this document, their name should follow the name of their supervisor.

<odesi> format

Formatting Note:
- Format for Agency/Institution and Place of Work/Department:
    Name of Agency/Institution. Name of Place of Work/Department
        (yes, that is a period between the two place names!)
- Format for Names:
    Last name, First name

Example 1:
<rspStmt>
<AuthEnty affiliation="**University of Guelph. Data Resource Centre**">**Perry, Carol**</AuthEnty>
<AuthEnty affiliation="**University of Guelph. Data Resource Centre**"> **Edwards, Michelle**</AuthEnty>
<AuthEnty affiliation="**University of Guelph. Data Resource Centre**"> **StudentLastName, StudentFirstName**</AuthEnty> </rspStmt>

Example 2:
    <rspStmt>
    <AuthEnty affiliation="**Queen's University. Canadian Opinion Research
        Archives**"></AuthEnty>
    <AuthEnty affiliation="**Queen's University. Canadian Opinion Research Archives**">
        Moon, Anne-Marie</AuthEnty>
     </rspStmt>

1.1.3   <prodStmt>   Production statement

- Optional
- Not Repeatable
- Attributes: ID, xml:lang, source

Description: Production statement for the marked-up document. Contains the following
elements:

1.1.3.1   <producer>   Producer

- Optional
- Repeatable
- Attributes: ID, xml:lang, source, abbr, affiliation, role

Description: The producer of the marked-up document is the person or organization with the
financial or administrative responsibility for the physical processes whereby the marked-up
document was brought into existence. Equivalent to Dublin Core Publisher.

Note:
     If there is more than one producer of this document, they should all be listed. See
     Example 2.

Example 1:
    <prodStmt> <producer abbr="**DRC**" affiliation="**University of Guelph**">**Data Resource
     Centre** </producer>

Example 2:
    <prodStmt> <producer abbr="**DC**" affiliation="**Carleton University**"> **Data Centre**
    </producer>
    <producer abbr="**SSDC**" affiliation="**Queen's University**">**Social Science Data Centre**
    </producer>

1.1.3.2   <copyright>   Copyright

- Optional
- Not Repeatable
- Attributes: ID, xml:lang, source

Description: Copyright statement for the marked-up document. Equivalent to Dublin Core
Rights.

Note:
   If there is more than one copyright listed here, they should be put in order of the most
   recent one to the oldest one.

Example 1:
   <copyright> **Copyright © University of Guelph. Data Resource Centre, 2004**
   </copyright>

Example 2:
   <copyright> **Copyright © Carleton University. Data Centre, 2008, 2006** </copyright>
   Explanation of Example 2:
   This file was originally produced by Carleton University in 2006 and then was updated
   by Carleton University in 2008.

Example 3:
   <copyright> **Copyright © Carleton University. Data Centre, 2008** </copyright>
   <copyright> **Copyright © University of Guelph. Data Resource Centre, 2004**
   </copyright>

1.1.3.3   <prodDate>   Production Date

- Optional
- Repeatable
- Attributes: <u>ID, xml:lang, source</u>, date

Description: Date the marked-up document was produced (not distributed or archived). The
ISO standard for dates (YYYY-MM-DD) is recommended for use with the date attribute.
Equivalent to Dublin Core Date.

Note:
   If there is more than one production date for this document, they should all be listed,
   starting with the most recent one. (See Example 2.)

Example 1:
   <prodDate date="**2004-06-07**"> **June 7, 2004**</prodDate>

Example 2:
   <prodDate date="**2008-01-17**">**2008-01-17**</prodDate>
   <prodDate date="**2007-11-16**">**2007-11-16**</prodDate>

1.1.3.4   <prodPlac>   Place of Production

- Optional
- Repeatable
- Attributes: <u>ID, xml:lang, source</u>

Description: Address of the archive or organization that produced the work.

<odesi> format

Formatting Note 1:
   No abbreviations should be used.

<odesi> format

Formatting Note 2:
   University/Institution, City, Province/State, Country

<odesi> format

Formatting Note 3:
> If the name of the Department/Agency is needed for clarification, include it after the name of the country with a ":" as a separator. (See Examples 4 and 5)

Example 1:
> <prodPlac>**University of Guelph, Guelph, Ontario, Canada**</prodPlac>

Example 2:
> <prodPlac>**Carleton University, Ottawa, Ontario, Canada**</prodPlac>

Example 3:
> <prodPlac> **Inter-university Consortium for Political and Social Research, Ann Arbor, Michigan, United States** </prodPlac>

Example 4:
> <prodPlac>**Queen's University, Kingston, Ontario, Canada: Canadian Opinion Research Archives**</prodPlac>

Example 5:
> <prodPlac>**University of Guelph, Guelph, Ontario, Canada**</prodPlac>
> <prodPlac>**Ottawa, Ontario, Canada: Statistics Canada. Special Surveys Division**</prodPlac>

1.1.3.5   <software>   Software used in Production

- Optional
- Repeatable
- Attributes: <u>ID, xml:lang, source</u>, date, version

Description: Software used to produce the work. A "version" attribute permits specification of the software version number. The "date" attribute is provided to enable specification of the date (if any) for the software release. The ISO standard for dates (YYYY-MM-DD) is recommended for use with the date attribute.

Note:
> See Appendix B for more information on this tag.

Example 1:
> <software version="**9.1.3**">**SAS**</software>

Example 2:
> <software version="**14.0**">**SPSS**</software>

1.1.3.6   <fundAg>   Funding Agency/Sponsor

- Optional
- Repeatable
- Attributes: <u>ID, xml:lang, source</u>, abbr, role

Description: The source(s) of funds for production of the work. If different funding agencies sponsored different stages of the production process, use the "role" attribute to distinguish them.

Note 1:
    See Appendix E for more information on this tag.

Note 2:
    Use only if there is external funding for the production of this document that is outside of normal operations

Note 2:
    Includes only the name of the agency. Does NOT include the location or role of the agency.

Example 1:
    <fundAg abbr="" role="infrastructure">**OntarioBuys**</fundAg>

Example 2:
    <fundAg abbr="**OCUL**" role="infrastructure">**Ontario Council of University Libraries**</fundAg>
    <fundAg abbr="**CU**" role="infrastructure">**Carleton University. Data Centre**</fundAg>

Example 3:
    <fundAg abbr="**NSF**" role="infrastructure">**National Science Foundation**</fundAg>

1.1.3.7   <grantNo>   Grant Number

- Optional
- Repeatable
- Attributes: <u>ID, xml:lang, source</u>, agency, role

Description: The grant/contract number of the project that sponsored the effort. If more than one, indicate the appropriate agency using the "agency" attribute. If different funding agencies sponsored different stages of the production process, use the "role" attribute to distinguish the grant numbers.

Example:
    <grantNo agency="**Bureau of Justice Statistics**">**J-LEAA-018-77**</grantNo>

1.1.4   <distStmt>   Distributor statement

- Optional
- Not Repeatable
- Attributes: <u>ID, xml:lang, source</u>

Description: Distribution statement for the marked-up document

1.1.4.1   <distrbtr>   Distributor

- Optional
- Repeatable
- Attributes: <u>ID, xml:lang, source</u>, abbr, affiliation, URI

Description: The organization designated by the author or producer to generate copies of particular marked-up documentation including any necessary editions or revisions. Names and addresses may be specified and other archives may be co-distributors. A URI attribute is included to provide an URN or URL to the ordering service or download facility on a website.

Example 1:
<distrbtr abbr="**DRC**" affiliation="**University of Guelph**" URI="**http://tdr-tug-libraries.on.ca**">**Guelph, Ontario: University of Guelph. Data Resource Centre**</distrbtr>

Example 2:
<distrbtr abbr="**SSDC**" affiliation="**Queen's University**" URI="**http://library.queensu.ca/webdoc/ssdc/**"> **Kingston, Ontario: Queen's University. Social Science Data Centre**</distrbtr>

Example 3:
<titl>**Canadian Gallup Poll, May 1949, #186**</titl>
distrbtr abbr="**DC**" affiliation="**Carleton University**" URI="**http://www.library.carleton.ca/ssdata/surveys/pop_gallup.html**">**Ottawa, Ontario: Carleton University. Data Centre**</distrbtr>

1.1.4.2   <contact>   Contact Persons

- Optional
- Repeatable
- Attributes: <u>ID, xml:lang, source</u>, affiliation, URI, email

Description: Names and addresses of institutions responsible for the work. The URI attribute should be used to indicate a URN or URL for the homepage of the contact institution. The email attribute is used to indicate an email address for the contact institution.

Example 1:
<contact affiliation="**University of Guelph**" email=" **drc@listserv.uoguelph.ca** ">**Data Resource Centre**</contact>

Example 2:
<contact affiliation="**Carleton University**" mail="**ssdata@ccs.carleton.ca**">**Data Centre**</contact>

Example 3:
<contact affiliation="**Queen's University**" mail="**madgic@queensu.ca**">**Social Science Data Centre**</contact>

1.1.5   <serStmt>   Series Statement

- Optional
- Not Repeatable
- Attributes: <u>ID, xml:lang, source</u>, URI

Description: Series statement for the marked-up document. The URI attribute is provided to point to a central Internet repository of series information.

1.1.5.1   \<serName\>   Series Name

- Optional
- Repeatable
- Attributes: <u>ID, xml:lang, source</u>, abbr

Description: The name of the series to which the marked-up document belongs. This will probably be the same as the Series Name for the study or data collection (2.1.5.1).

Example 1:
     \<serName\>**Household Internet Use Survey**\</serName\>

Example 2:
     \<serName\>**General Social Survey**\</serName\>

Example 3:
     \<serName\>**Canadian Tobacco Use Monitoring Survey**\</serName\>

Example 4:
     \<serName\>**Canadian Community Health Survey**\</serName\>

1.1.5.2   \<serInfo\>   Series Information

- Optional
- Repeatable
- Attributes: <u>ID, xml:lang, source</u>

Description: Contains a history of the series and a summary of those features that apply to the series as a whole.

Note 1:
     Include the starting date of the series.

Note 2:
     If the name of the series changes, this should be included in the description.

Note 3:
     See Appendix C for more information on this tag.

Example 1:
     \<serInfo\>**Conducted since 1997, the Survey of Household Spending provides Information about the spending habits, dwelling characteristics and household equipment of Canadian households. Annual from 1997 – 1999, it has become a biennial survey as of 2000**.\</serInfo\>

Example 2:
     \<serInfo\>**The Canadian Tobacco Use Monitoring Survey was conducted for Health Canada to provide data on tobacco use and related issues. The primary objective of the survey is to track changes in smoking status, especially for population most at risk, such as the 15-24 years olds. The survey will allow Health Canada to estimate smoking prevalence for age groups 15-24 and 25+ by province and by gender on a semi-annual basis.** \</serInfo\>

1.1.6   <verStmt>   Version Statement

- Optional
- Repeatable
- Attributes: <u>ID, xml:lang, source</u>

Description: Version statement for the marked-up document

1.1.6.1   <version>   Version

- Optional
- Not Repeatable
- Attributes: <u>ID, xml:lang, source</u>, date, type

Description: Also known as release or edition. If there have been substantive changes in the marked-up document since its creation, this statement should be used. The ISO standard for dates (YYYY-MM-DD) is recommended for use with the date attribute.

Note 1:
   Every time this document is changed, this tag should be used, with the most recent version date listed first, followed by the older one(s).

Note 2:
   Every time there is a new version of this document, the version date is used in place of a version number to avoid confusion in version numbering.

Example 1:
   <version date="2007-11-10">
   Version January 18, 2008
   Version November 10, 2007 edited from Data Resource Centre, University of Guelph
   </version>

Example 2:
   <version type="**edition**" date="**1999-01-25**">**January 25, 1999**</version>

Example 3:
   <version type="**version**" date="**2000-01-01**">**January 1, 2000**</version>

1.1.6.2   <verResp>   Version Responsibility Statement

- Optional
- Not Repeatable
- Attributes: <u>ID, xml:lang, source</u>, affiliation

Description: The organization responsible for the version of the work.

Note 1:
   Every time this document is changed, this tag should be used.

Note 2:
   Every time there is a new version of this document, the version date is used in place of a version number to avoid confusion in version numbering.

Example 1:
　　<verResp>**University of Guelph. Data Resource Centre**</verResp>

Example 2:
　　<verResp>**Inter-university Consortium for Political and Social Research**</verResp>

1.1.6.3　　　　　　<notes>　　　　　　Notes and Comments

- Optional
- Repeatable
- Attributes: <u>ID, xml:lang, source</u>, type, subject, level, resp, sdatrefs

Description: Used to indicate additional information regarding the version or the version responsibility statement for the marked-up document, in particular to indicate what makes a new version different from its predecessor. "Notes" sections appear in several places in the DTD. The attributes for notes permit a controlled vocabulary to be developed (type and subject), the level of the DTD to which the note refers to be identified (study, file, variable, etc.), and the author of the note to be indicated (responsibility).

Note 1:
　　Every time this document is changed, this tag should be used, with the most recent note being entered first, followed by the older notes.

Note 2:
　　Every time there is a new version of this document, the version date is used in place of a version number to avoid confusion in version numbering.

Example 1:
　　<notes>**Additional study information was added to this document**.</notes>

Example 2:
　　<notes resp="**Smith, Jane**">**Additional information on derived variables has been added to this marked-up version of the documentation**.</notes>

Example 3:
　　<notes> **Version January 18, 2008 - made file compliant to <odesi> Best Practices Standards; added documentation for each variable.**
　　Version November 10, 2007 - changed information in Document Description, and Other Materials.
　　</notes>

1.1.7　　　　<biblCit>　　　　　　　　Bibiliographic Citation

- Optional
- Not Repeatable
- Attributes: <u>ID, xml:lang, source</u>, format

Description: Complete bibliographic reference containing all of the standard elements of a citation that can be used to cite the work. The "format" attribute is provided to enable specification of the particular citation style used, e.g., APA, MLA, Chicago, etc.

Note 1:
Remember that this tag is for the bibliographic citation for this particular document and it usually will not be the same as the bibliographic citation for the actual study (tag 2.1.7).

Note 2:
If you have a student doing the mark-up, their name is not included in this tag as it is included elsewhere in this document.

Example 1:
<biblCit format="APA"> **Fry, Jane, 2007. Gallup Poll, May 1945, #142 [Canada]. [DDI Codebook file]. Data Centre. Carleton University, Ottawa, Ontario: [producer and distributor].** </biblCit>

Example2 :
<biblCit format="APA"> **Perry, C.M. and Edwards, A.M. 2005. Survey of Household Spending 2003 [Canada]. [DDI Codebook file]. Data Resource Centre. University of Guelph, Guelph, Ontario: [producer and distributor].** </biblCit>

1.1.8   <holdings>   Holdings Information

- Optional
- Repeatable
- Attributes: <u>ID, xml:lang, source</u>, location, callno, URI, media

Description: Information concerning either the physical or electronic holdings of the cited work. Attributes include: location--The physical location where a copy is held; callno--The call number for a work at the location specified; and URI--A URN or URL for accessing the electronic copy of the cited work.

Example:
<holdings location="**University of Guelph. Data Resource Centre**" URI="**http://nesstar.tdr.uoguelph.ca/CONSUM/SHS/2002/DOCS/**">**DRC XML Codebook**</holdings>

1.2   <docStatus>   docStatus

- Optional
- Repeatable
- Attributes: <u>ID, xml:lang, source</u>

Description: Use this field to indicate if the documentation is being presented/distributed before it has been finalized. Some data producers and social science data archives employ data processing strategies that provide for release of data and documentation at various stages of processing.

Example:
<docStatus>**This marked-up document includes a provisional data dictionary and brief citation only for the purpose of providing basic access to the data file. A complete codebook will be published at a later date.**</docStatus>

SECTION 2.0   STUDY DESCRIPTION   <stdyDscr>

The Study Description consists of information about the data collection, study, or compilation that the DDI-compliant documentation file describes. This section includes information about how the study should be cited, who collected or compiled the data, who distributes the data, keywords about the content of the data, summary (abstract) of the content of the data, data collection methods and processing, etc. Note that some content of the Study Description's Citation -- e.g., Responsibility Statement -- may be identical to that of the Documentation Citation. This is usually the case when the producer of a data collection also produced the print or electronic codebook for that data collection.

Tags and corresponding DTD numbers are from DDI Version 2.0 codebook.

Source: DDI Codebook
http://www.icpsr.umich.edu/DDI/dtd/version2-1-all.html
http://www.icpsr.umich.edu/DDI/dtd/version2-1-tree.html

DTD Numbers	Tags
2.0	<stdyDscr>
2.1	<citation>
2.1.1	<titlStmt>
2.1.1.1	<titl>
2.1.1.2	<subTitl>
2.1.1.3	<altTitl>
2.1.1.4	<parTitl>
2.1.1.5	<IDNo>
2.1.2	<rspStmt>
2.1.2.1	<AuthEnty>
2.1.2.2	<othID>
2.1.3	<prodStmt>
2.1.3.1	<producer>
2.1.3.2	<copyright>
2.1.3.3	<prodDate>
2.1.3.4	<prodPlac>
2.1.3.5	<software>
2.1.3.6	<fundAg>
2.1.3.7	<grantID>
2.1.4	<distStmt>
2.1.4.1	<distrbtr>
2.1.4.2	<contact>
2.1.4.3	<depositr>
2.1.4.4	<depDate>
2.1.4.5	<distDate>
2.1.5	<serStmt>
2.1.5.1	<serName>
2.1.5.2	<serInfo>
2.1.6	<verStmt>
2.1.6.1	<version>
2.1.6.2	<verResp>

2.1.6.3	<notes>
2.1.7	<biblCit>
2.1.9	<notes>
2.2	<stdyInfo>
2.2.1	<subject>
2.2.1.1	<keyword>
2.2.1.2	<topcClas>
2.2.2	
2.2.3	<sumDscr>
2.2.3.1	<timePrd>
2.2.3.2	<collDate>
2.2.3.3	<nation>
2.2.3.4	<geogCover>
2.2.3.5	<geogUnit>
2.2.3.8	<anlyUnit>
2.2.3.9	<universe>
2.2.3.10	<dataKind>
2.2.4	<notes>
2.3	<method>
2.3.1	<dataColl>
2.3.1.1	<timeMeth>
2.3.1.2	<dataCollector>
2.3.1.3	<frequenc>
2.3.1.4	<sampProc>
2.3.1.6	<collMode>
2.3.1.7	<resInstru>
2.3.1.8	<sources>
2.3.1.8.1	<dataSrc>
2.3.1.8.2	<srcOrig>
2.3.1.9	<collSitu>
2.3.1.12	<weight>
2.3.2	<notes>
2.3.3	<anlyInfo>
2.3.3.1	<respRate>
2.3.3.2	<EstSmpErr>
2.4	<dataAccs>
2.4.1	<setAvail>
2.4.1.1	<accsPlac>
2.4.1.2	<origArch>
2.4.1.3	<avlStatus>
2.4.1.4	<collSize>
2.4.1.5	<complete>
2.4.1.6	<fileQnty>
2.4.1.7	<notes>
2.4.2	<useStmt>

2.4.2.3	<restrctn>
2.4.2.4	<contact>
2.4.2.5	<citReq>
2.4.2.7	<conditions>
2.4.2.8	<disclaimer>
2.5	<othrStdyMat>
2.5.1	<relMat>
2.5.2	<relStdy>
2.5.3	<relPubl>

Description of Tags and Working Examples

2.0   <stdyDscr>   Study Description

- Mandatory
- Repeatable
- Attributes: ID, xml:lang, source, access

Description: The Study Description consists of information about the data collection, study, or compilation that the DDI-compliant documentation file describes. This section includes information about how the study should be cited, who collected or compiled the data, who distributes the data, keywords about the content of the data, summary (abstract) of the content of the data, data collection methods and processing, etc. Note that some content of the Study Description's Citation -- e.g., Responsibility Statement -- may be identical to that of the Documentation Citation. This is usually the case when the producer of a data collection also produced the print or electronic codebook for that data collection

2.1   <citation>   Citation

- Mandatory
- Repeatable
- Attributes: ID, xml:lang, source, MARCURI

Description: Citation for the data collection described by the marked-up documentation. This element encodes the bibliographic information describing the data collection, including title information, statement of responsibility, production and distribution information, series and version information, text of a preferred bibliographic citation, and notes (if any). A MARCURI attribute is provided to link to the MARC record for this citation.

2.1.1   <titlStmt>   Title Statement

- Mandatory
- Not Repeatable
- Attributes: ID, xml:lang, source

Description: Title statement for the work at the appropriate level: marked-up document; marked-up document source; study; study description; and other materials for the study.

2.1.1.1   <titl>   Title

- Mandatory
- Not Repeatable
- Attributes: ID, xml:lang, source

Description: Contains the full authoritative title of the data collection. The data collection title will in most cases be identical to the title for the marked-up document (1.1.1.1) and the source document (1.4.1.1). A full title should indicate the geographic scope of the data collection as well as the time period covered.

\<odesi\> *Formatting Note 1:*
If the country name is in the title of the survey, it does not need to be put in square brackets at the end of the title statement.

\<odesi\> *Formatting Note 2:*
No abbreviations or short forms should be used in the title, eg. name of province, country.

\<odesi\> *Formatting Note 3:*
If the files are not in microdata format, i.e., Beyond 20/20 or Excel, this should be included after the title in square brackets. (See Example 8)

\<odesi\> *Formatting Note 4:*
If the files are only for the use of your institution, this should be included after the title in square brackets. (See Example 9)

Example 1:
<titl>**Sun Exposure Survey, 1996 [Canada]**</titl>

Example 2:
<titl>**Household Internet Use, 2001 [Canada]**</titl>

Example 3:
<titl>**General Social Survey, 2005 [Canada]: Cycle 19, Time Use, Main File**</titl>

Example 4:
<titl>**Canadian Tobacco Use Monitoring Survey, 2004: Cycle 1, Household File**</titl>

Example 5:
<titl>**Canadian Community Health Survey, 2005: Cycle 3.1, Health Sample File**</titl>

Example 6:
<titl>**Canadian Gallup Poll, May 1949, #186**</titl>

Example 7:
<titl>**Domestic Violence Experience in Omaha, Nebraska, 1986-1987 [United States]**</titl>

Example 8:
<titl>**Methods Used to Commit Homicide Offences in Canada, 1974 to 2003 [Beyond 20/20]**</titl>

Example 9:
<titl>**Australian Election Study, 1987 [In-House Use Only]**</titl>

2.1.1.2   <subTitl>   Subtitle

- Optional
- Repeatable
- Attributes: ID, xml:lang, source

Description: A subtitle is a secondary title used to amplify or state certain limitations of the main title. It may repeat information already in the main title.

Note:
   The subtitle includes everything after the colon in the title.

Example 1:
   <titl>**General Social Survey, 2005 [Canada]: Cycle 19, Time Use, Main File** </titl>
   <subTitl> **Cycle 19, Time Use, Main File**</subTitl>

Example 2:
   <titl>**Canadian Tobacco Use Monitoring Survey, 2004: Cycle 1, Household File** </titl>
   <subTitl> **Cycle 1, Household File**</subTitl>

Example 3:
   <titl>**Canadian Community Health Survey, 2005: Cycle 3.1, Sub-sample 1: Health Utility Index (HUI), Fruit and Vegetable Consumption and Labour Force (long form)** </titl>
   <subTitl>**Cycle 3.1, Sub-sample 1: Health Utility Index (HUI), Fruit and Vegetable Consumption and Labour Force (long form)**</subTitl>

2.1.1.3   <altTitl>   Alternate Title

- Optional
- Repeatable
- Attributes: ID, xml:lang, source

Description: The alternative title may be the title by which a data collection is commonly referred to or it may be an abbreviation for the title.

<odesi> format

Formatting Note:
   If there is a subtitle after the title, it should be included in the alternate title.

Example 1:
   <titl>**Sun Exposure Survey, 1996 [Canada]**</titl>
   <altTitl>**SES 1996**</altTitl>

Example 2:
    <titl>**Household Internet Use, 2001 [Canada]**</titl>
   <altTitl>**HIUS 2001**</altTitl>

Example 3:
   <titl>**General Social Survey, 2005 [Canada]: Cycle 19, Time Use, Main File** </titl>
   <altTitl>**GSS19 2005: Cycle 19, Time Use, Main File**</altTitl>

Example 4:
   <titl>**Canadian Tobacco Use Monitoring Survey, 2004: Cycle 1, Household File** </titl>
   <altTitl>**CTUMS 2005: Cycle 1, Household File**</altTitl>

Example 5:
   <titl>**Canadian Community Health Survey, 2005: Cycle 3.1, Health Sample File**</titl>
   <altTitl>**CCHS 2005: Cycle 3.1, Health Sample File**</altTitl>

2.1.1.4   <parTitl>   Parallel Title

- Optional
- Repeatable
- Attributes: <u>ID, xml:lang, source</u>

Description: Title translated into another language.

Example:
   <titl>**Survey of Household Spending, 2002 [Canada]**</titl>
   <parTitl>**Enquête sur les dépenses des ménages, 2002 [Canada]**</parTitl>

2.1.1.5   <IDNo>   Identification Number

- Optional
- Repeatable
- Attributes: <u>ID, xml:lang, source</u>, agency, level

Description: Unique string or number (producer's or archive's number) for the data collection. An "agency" attribute is supplied.

Note 1:
   This ID number is the same for the document description and the study description, ie. 2.1.1.5 is the same as 1.1.1.5.

Note 2:
   For Statistics Canada surveys, the catalogue number refers to the microdata file.

Formatting Note 1:
- Language: E = English, F = French
- Year: yyyy or yyyy-mm-dd
- Lowercase: should be used for everything except the catalogue number and the language abbreviation
- Surveys with cycle numbers and sub-numbers: use a dash between the numbers and not a period, eg. for cycle 2.1 the format would be c2-1
- ICPSR surveys: use their ID# and add the short form if there is a subset, eg. icpsr9721im (see example 6)

Formatting Note 2:
   This is the format to be used:
      acronym_CatalogueNumber_language_year_subset

Example 1:
   <titl>**Sun Exposure Survey, 1996 [Canada]**</titl>
   <IDNo="DRC">**ses_82M0019_E_1996**</IDNo>

Example 2:
   <titl>**General Social Survey, 2005 [Canada]: Cycle 19, Time Use, Main File**</titl>
   <IDNo>**gss_12M0019_E_2005_c19_main-file**</IDNo>

Example 3:
   <titl>**Canadian Tobacco Use Monitoring Survey, 2004: Cycle 1, Household File**</titl>
   <IDNo="DRC">**ctums_82M0020_E_2004_c1_household-file**</IDNo>

Example 4:
    <titl>**Canadian Community Health Survey, 2005: Cycle 3.1, Health Sample File**</titl>
    <IDNo="DRC">**cchs_82M0013_E_2005_c3-1_health-sample**</IDNo>

Example 5:
    <titl>**Canadian Gallup Poll, May 1949, #186**</titl>
    <IDNo="DC">**cipo_186_E_1949-05**</IDNo>

Example 6:
    <titl>**Descriptors and Measurements of the Height of Runaway Slaves and Indentured Servants in the United States: Runaway Indentured Servant Height Measurements, 1700-1850**</titl>
    <IDNo>**icpsr9721im**</IDNo>

2.1.2   <rspStmt>   Responsibility Statement

- Optional
- Not Repeatable
- Attributes: <u>ID, xml:lang, source</u>

Description: Responsibility for the data collection. Responsibility for the creation of the work at the appropriate level: marked-up document; marked-up document source; study; study description, other material; other material for study.

2.1.2.1   <AuthEnty>   Authoring Entity/ Primary Investigator

- Optional
- Repeatable
- Attributes: <u>ID, xml:lang, source</u>, affiliation

Description: The person, corporate body, or agency responsible for the data collection's substantive and intellectual content. Repeat the element for each author, and use the affiliation attribute if available. Invert first and last name and use commas.

Note 1:
    The author entity is based on information from the User's Guide, usually on the title page.

Note 2:
    If a subdivision of the parent agency/institution exists, it should be included here.

<odesi>
  format

Formatting Note 1:
    This is the format to be used:
    Name of Institution/University. Name of Data Centre/Division

Example 1:
    <AuthEnty>**Statistics Canada. Income Statistics Division**</AuthEnty>

Example 2:
    <AuthEnty>**United Nations. Statistics Division**</AuthEnty>

Example 3:
    <AuthEnty>**Gallup Canada Inc.**</AuthEnty>

2.1.2.2   <othId>   Other Identifications /Acknowledgments

- Optional
- Repeatable
- Attributes: <u>ID, xml:lang, source</u>, type, role, affiliation

Description: Statements of responsibility not recorded in the title and statement of responsibility areas. Indicate here the persons or bodies connected with the work, or significant persons or bodies connected with previous editions and not already named in the description. For example, the name of the person who edited the marked-up documentation might be cited in 1.1.2.2, using the "role" and "affiliation" attributes. Other identifications/ acknowledgements for data collection (2.1.2.2) map to Dublin Core Contributor element.

Note:
     See Appendix C for more information on this tag.

Example:
     <titl>**Canadian Tobacco Use Monitoring Survey, 2004: Cycle 1, Household File**</titl>
     <othId><p>"**Health Canada**"></p></othId>

2.1.3   <prodStmt>   Production Statement

- Optional
- Not Repeatable
- Attributes: <u>ID, xml:lang, source</u>

Description: Production statement for the work at the appropriate level: marked-up document; marked-up document source; study; study description, other material; other material for study.

2.1.3.1   <producer>   Producer

- Optional
- Repeatable
- Attributes: <u>ID, xml:lang, source</u>, abbr, affiliation, role

Description: The producer of the data collection is the person or organization with the financial or administrative responsibility for the physical processes whereby the data collection was brought into existence. Use the role attribute to distinguish different stages of involvement in the production process, such as original producer.

Example 1:
     <producer>**Gallup Canada Inc.**</producer>

Example 2:
     <producer abbr="**SSD**" affiliation="**Special Surveys Division**">**Statistics Canada**</producer>

Example 3:
     <producer abbr="**LSD**" affiliation="**Labour Statistics Division**">**Statistics Canada**</producer>

Example 4:
<producer abbr="**ICPSR**" affiliation**="Institute for Social Research**">**Inter-university Consortium for Political and Social Research**</producer>

2.1.3.2 <copyright> Copyright

- Optional
- Not Repeatable
- Attributes: ID, xml:lang, source

Description: Copyright statement for the data collection. Copyright for data collection (2.1.3.2) maps to Dublin Core Rights. Inclusion of this element is recommended element.

Note 1:
The date used is the actual release date of the document, which is not necessarily the same as the date the survey was conducted.

Note 2:
See Appendix C for more information on this tag.

Example 1:
<titl>**Canadian Gallup Poll, May 1949, #186**</titl>
<copyright>**Copyright © Gallup Canada Inc., 1950**</copyright>

Example 2:
<titl>**Canadian Tobacco Use Monitoring Survey, 2004: Cycle 1, Household File**</titl>
<copyright>**Copyright © Statistics Canada, 2005**</copyright>

2.1.3.3 <prodDate> Date of Production

- Optional
- Repeatable
- Attributes: ID, xml:lang, source, date

Description: Date the data collection was produced (not distributed or archived). The ISO standard for dates (YYYY-MM-DD) is recommended for use with the date attribute. Production date for data collection (2.1.3.3) maps to Dublin Core Date element.

Note 1:
This is the release date of the data.

Note 2:
When there is more than one date, the most recent one is listed first.

Example 1:
<prodDate date="2003-11-30">**November 30, 2003**</prodDate>

Example 2:
<titl>**Canadian Tobacco Use Monitoring Survey, 2004: Cycle 1, Household File**</titl>
<prodDate date="2005">**2005**</prodDate>

2.1.3.4   <prodPlac>   Place of Production

- Optional
- Repeatable
- Attributes: <u>ID, xml:lang, source</u>

Description: Address of the archive or organization that produced the work.

Formatting Note 1:
     No abbreviations should be used.

Formatting Note 2:
     This is the format to be used:
         Name of Institution/University, City, Province/State, Country

Formatting Note 3:
     If the name of the Department/Agency is needed for clarification, include it after the name of the country with a ":" as a separator. (See Examples 4 and 5)

Example 1:
     <prodPlac>**University of Guelph, Guelph, Ontario, Canada**</prodPlac>

Example 2:
     <prodPlac>**Carleton University, Ottawa, Ontario, Canada**</prodPlac>

Example 3:
     <prodPlac> **Inter-university Consortium for Political and Social Research, Ann Arbor, Michigan, United States** </prodPlac>

Example 4:
     <prodPlac>**Queen's University, Kingston, Ontario, Canada: Canadian Opinion Research Archives**</prodPlac>

Example 5:
     <prodPlac>**Statistics Canada, Ottawa, Ontario, Canada: Special Surveys Division**</prodPlac>

2.1.3.5   <software>   Software used in Production

- Optional
- Repeatable
- Attributes: <u>ID, xml:lang, source</u>, date, version

Description: Software used to produce the work. A "version" attribute permits specification of the software version number. The "date" attribute is provided to enable specification of the date (if any) for the software release. The ISO standard for dates (YYYY-MM-DD) is recommended for use with the date attribute.

Note:
     See Appendix B for more information on this tag.

Example 1:
   <software version="**9.1.3**">**SAS**</software>

Example 2:
   <software version="**14.0**">**SPSS**</software>

2.1.3.6   <fundAg>   Funding Agency/Sponsor

- Optional
- Repeatable
- Attributes: <u>ID, xml:lang, source</u>, abbr, role

Description: The source(s) of funds for production of the work. If different funding agencies sponsored different stages of the production process, use the "role" attribute to distinguish them.

Note 1:
   This tag is used if there was external funding for the survey that is outside of normal operations.

Note 2:
   Includes only the name of the agency. Does NOT include the location or role of the agency.

Example 1:
   <fundAg abbr="**CRIC**" role="**infrastructure**">**Centre for Research and Information on Canada**</fundAg>

Example 2:
   <fundAg abbr="**CRIC**" role="**infrastructure**">**Centre for Research and Information on Canada**</fundAg>
   <fundAg abbr="**OCUL**" role="**infrastructure**">**Ontario Council of University Libraries**</fundAg>

Example 3:
   <fundAg abbr="**NSF**" role="**infrastructure**">**National Science Foundation**</fundAg>

2.1.3.7   <grantNo>   Grant Number

- Optional
- Repeatable
- Attributes: <u>ID, xml:lang, source</u>, agency, role

Description: The grant/contract number of the project that sponsored the effort. If more than one, indicate the appropriate agency using the "agency" attribute. If different funding agencies sponsored different stages of the production process, use the "role" attribute to distinguish the grant numbers.

Example:
   <grantNo agency="**Bureau of Justice Statistics**">**J-LEAA-018-77**</grantNo>

2.1.4   \<distStmt\>   Distributor Statement

- Optional
- Not Repeatable
- Attributes: <u>ID, xml:lang, source</u>

Description: Distribution statement for the data collection.

2.1.4.1   \<distrbtr\>   Distributor

- Optional
- Repeatable
- Attributes: <u>ID, xml:lang, source</u>, abbr, affiliation, URI

Description: The organization designated by the author or producer to generate copies of a particular data collection including any necessary editions or revisions. Names and addresses may be specified, and other archives may be co-distributors. A URI attribute is included to provide an URN or URL to the ordering service or download facility on a website.

 Note:
   If there is more than one affiliation/organization that is a distributor, this tag is repeated.

 Example 1:
   \<titl\>**Canadian Community Health Survey, 2005: Cycle 3.1, Health Sample File**\</titl\>
   \<distrbtr abbr="**DLI**" affiliation="**Statistics Canada**"
   URI="**http://www.statcan.ca/english/Dli/dli.htm**"\>**Data Liberation Initiative**\</distrbtr\>

 Example 2:
    \<titl\>**Canadian Gallup Poll, May 1949, #186**\</titl\>
   \<distrbtr abbr="**DC**" affiliation="**Carleton University**"
   URI="**http://www.library.carleton.ca/ssdata/surveys/pop_gallup.html**"\>**Data Centre**\</distrbtr\>

 Example 3:
   \<distrbtr abbr="**CORA**" affiliation="**Queen's University**"
   URI="**http://www.queensu.ca/cora/**"\>**Canadian Opinion Research Archive**\</distrbtr\>

 Example 4:
   \<distrbtr abbr="**DLI**" affiliation="**Statistics Canada**"
   URI="**http://www.statcan.ca/english/Dli/dli.htm**"\>**Data Liberation Initiative**\</distrbtr\>
   \<distrbtr abbr="**CORA**" affiliation="**Queen's University**"
   URI="**http://www.queensu.ca/cora/**"\>**Canadian Opinion Research Archive**\</distrbtr\>

2.1.4.2　　　　<contact>　　　　　Contact Persons

- Optional
- Repeatable
- Attributes: <u>ID, xml:lang, source</u>, affiliation, URI, email

Description: Names and addresses of individuals responsible for the work. Individuals listed as contact persons will be used as resource persons regarding problems or questions raised by the user community. The URI attribute should be used to indicate a URN or URL for the homepage of the contact individual. The email attribute is used to indicate an email address for the contact individual.

Example 1:
　　<contact affiliation="**University of Guelph**" email="**drc@listserv.uoguelph.ca**"> **Data Resource Centre** </contact>

Example 2:
　　<contact affiliation="**Statistics Canada**"
　　uri="**http://www.statcan.ca/english/Dli/dli.htm**">**Data Liberation Initiative (DLI)**</contact>

2.1.4.3　　　　<depositr>　　　　　Depositor

- Optional
- Repeatable
- Attributes: <u>ID, xml:lang, source</u>, abbr, affiliation

Description: The name of the institution (and person, where applicable) who provided this work to the archive storing it.

Note:
　　This tag is not about the producer of the data, rather the secondary distributor, that is, your institution.

Example 1:
　　<depositr abbr="" affiliation="**Gallup Canada Inc.**">**Gallup Canada Inc.** </depositr>

Example 2:
　　<depositr abbr="" affiliation="**Carleton University**">**Bear, Dr. Smoke E./ Fire Hazard Prevention Project** </depositr>

2.1.4.4　　　　<depDate>　　　　　Date of Deposit

- Optional
- Repeatable
- Attributes: <u>ID, xml:lang, source</u>, date

Description: The date that the work was deposited with the archive that originally received it. The ISO standard for dates (YYYY-MM-DD) is recommended for use with the "date" attribute.

Note 1:
>   This tag is not about the producer of the data, rather the secondary distributor, that is, your institution.

Note 2:
>   When there is more than one date, the most recent one is listed first.

Example 1:
>   <depDate date="**1999-01-25**">**January 25, 1999**</depDate>

Example 2:
>   <depDate date="**2008-02-25**">**February 25, 2008**</depDate>
>   <depDate date="**1999-01-25**">**January 25, 1999**</depDate>

2.1.4.5   <distDate>   Date of Distribution

- Optional
- Not Repeatable
- Attributes: <u>ID, xml:lang, source</u>, date

Description: Date that the work was made available for distribution/presentation. The ISO standard for dates (YYYY-MM-DD) is recommended for use with the "date" attribute.

Note:
>   This tag is not about the producer of the data, rather the secondary distributor, that is, your institution.

Example:
>   <distDate date="**1999-01-25**">**January 25, 1999**</distDate>

2.1.5   <serStmt>   Series Statement

- Optional
- Not Repeatable
- Attributes: <u>ID, xml:lang, source</u>, URI

Description: Series statement for the data collection. The URI attribute is provided to point to a central Internet repository of series information.

2.1.5.1   <serName>   Name of Series

- Optional
- Repeatable
- Attributes: <u>ID, xml:lang, source</u>, abbr

Description: The name of the data series to which the collection belongs.

Note:
>   See Appendix C for more information on this tag.

Example 1:
   <serName>**Survey of Household Spending**</serName>

Example 2:
   <serName abbr="**SHS"**> **Survey of Household Spending** </serName>

Example 3:
   <serName>**Household Internet Use Survey**</serName>

 Example 4:
   <serName>**General Social Survey**</serName>

 Example 5:
   <serName>**Canadian Tobacco Use Monitoring Survey**</serName>

 Example 6:
   <serName>**Canadian Community Health Survey**</serName>

2.1.5.2   <serInfo>   Series Information

- Optional
- Repeatable
- Attributes: ID, xml:lang, source

Description: Contains a history of the data series and a summary of those features that apply
to the data series as a whole.

Note 1:
   Include the starting date of the series.

Note 2:
   If the name of the series changes, include this in the description.

Note 3:
   See Appendix C for more information on this tag.

Example 1:
   <serInfo>**Conducted since 1997, the Survey of Household Spending provides
   Information about the spending habits, dwelling characteristics and household
   equipment of Canadian households. Annual from 1997 – 1999, it has become a
   biennial survey as of 2000**.</serInfo>

Example 2:
   <serInfo>**The Canadian Tobacco Use Monitoring Survey was conducted for Health
   Canada to provide data on tobacco use and related issues. The primary objective
   of the survey is to track changes in smoking status, especially for population most
   at risk, such as the 15-24 years olds. The survey will allow Health Canada to
   estimate smoking prevalence for age groups 15-24 and 25+ by province and by
   gender on a semi-annual basis.** </serInfo>

2.1.6   <verStmt>   Version Statement

- Optional
- Repeatable
- Attributes: <u>ID, xml:lang, source</u>

Description: Version statement for the data collection

2.1.6.1   <version>   Version

- Optional
- Not Repeatable
- Attributes: <u>ID, xml:lang, source</u>, date, type

Description: Also known as release or edition. If there have been substantive changes in the data collection since its creation, this statement should be used. The ISO standard for dates (YYYY-MM-DD) is recommended for use with the date attribute.

Note:
   Every time there is a new version of this document, the version date is used in place of a version number to avoid confusion in version numbering.

Example 1:
   <version>**October 31, 2007**</version>

Example 2:
   <version type="**edition**" date"="**2002-12-15**">**December 15, 2002** </version>

2.1.6.2   <verResp>   Version Responsibility Statement

- Optional
- Not Repeatable
- Attributes: <u>ID, xml:lang, source</u>, affiliation

Description: The organization or person responsible for the version of the work.

Example 1:
   <verResp>**Inter-university Consortium for Political and Social Research**</verResp>

Example 2:
   <verResp>**Statistics Canada. Special Surveys Division**</verResp>

2.1.6.3   <notes>   Notes and Comments

- Optional
- Repeatable
- Attributes: <u>ID, xml:lang, source</u>, type, subject, level, resp, sdatrefs

Description: Used to indicate additional information regarding the version or the version responsibility statement for the data collection, in particular to indicate what makes a new version different from its predecessor. "Notes" sections appear in several places in the DTD. The attributes for notes permit a controlled vocabulary to be developed (type and subject), the level of the DTD to which the note refers to be identified (study, file, variable, etc.), and the author of the note to be indicated (resp).

Example:
<notes resp="**Doe, Jane**">**1997 Data files and documentation revised. The update standardizes the files with the 2000 file**.</notes>

2.1.7   <biblCit>   Bibliographic Citation

- Optional
- Not Repeatable
- Attributes: <u>ID, xml:lang, source</u>, format

Description: Complete bibliographic reference containing all of the standard elements of a citation that can be used to cite the data collection. The "format" attribute is provided to enable specification of the particular citation style used, e.g. APA, MLA, Chicago, etc.

Note:
Remember that this tag is for the bibliographic citation for this particular study and it probably will not be the same as the bibliographic citation for this actual document that is being marked up (tag 1.1.7).

<odesi>
format

Formatting Note :
Note: Cartographic Materials: A Manual of Interpretation for AACR2.
    The rules for unknown dates are as follows:
    [1971 or 1972] - one year or the other;
    [1969?] - probable date;
    [between 1906 and 1912] - used only for dates less than 20 years apart;
    [ca. 1960] - approximate date;
    [197-] - decade certain;
    [197-?] - probable decade;
    [18--] - century certain; and
    [18--?] - probable century.

Example 1:
<biblCit format="APA"> **Canadian Institute of Public Opinion (CIPO). 19--. Gallup Poll, May 1945, #142 [Canada]. [public-use microdata file]. Toronto, Ontario. Gallup Canada Inc. [distributor].**</biblCit>

Example 2:
<biblCit format="APA">**Statistics Canada. 2003. Survey of Household Spending [public-use microdata file]. Ottawa, Ontario: Statistics Canada. Income Statistics Division [producer and distributor].**</biblCit>

2.1.9   <notes>   Notes and comments

- Optional
- Repeatable
- Attributes: <u>ID, xml:lang, source</u>, type, subject, level, resp, sdatrefs

Description: For clarifying information/annotation regarding the parent element. The attributes for notes permit a controlled vocabulary to be developed ("type" and "subject"), indicate the "level" of the DDI to which the note applies (study, file, variable, etc.), and identify the author of the note ("resp").

Example 1:
    <stdyDscr><notes resp="**Smith, Jane**">**Data for 1998 have been added to this version of the data collection.**</notes></stdyDscr>

Example 2:
    <stdyInfo><notes>**Data on employment and income refer to the preceding year, although demographic data refer to the time of the survey.**</notes></stdyInfo>

2.2   <stdyInfo>   Study Scope

- Optional
- Repeatable
- Attributes: <u>ID, xml:lang, source</u>

Description: This section contains information about the data collection's scope across several dimensions, including substantive content, geography, and time.

2.2.1   <subject>   Subject

- Optional
- Not Repeatable
- Attributes: <u>ID, xml:lang, source</u>

Description: Subject information describing the data collection's intellectual content.

2.2.1.1   <keyword>   Keywords

- Optional
- Repeatable
- Attributes: <u>ID, xml:lang, source</u>, vocab, vocabURI

Description: Words or phrases that describe salient aspects of a data collection's content. Can be used for building keyword indexes and for classification and retrieval purposes. A controlled vocabulary can be employed. Maps to Dublin Core Subject. The vocab attribute is provided for specification of the controlled vocabulary in use, e.g., LCSH, MeSH, etc. The vocabURI attribute specifies the location for the full controlled vocabulary.

Note:
    See Appendix C for more information on this tag.

<odesi>
format

Formatting Note 1:
    If possible, this list should be alphabetized.

<odesi>
format

Formatting Note 2:
    Only the first letter of the first word in a Keywords phrase is capitalized, unless it is a proper name.

Example:
        <keyword>**Accommodation costs**</keyword>
        <keyword>**Educational attainment**</keyword>
        <keyword>**GST**</keyword>
        <keyword>**Lifestyle**</keyword>
        <keyword>**Physical activity**</keyword>
        <keyword>**Pierre Elliott Trudeau**</keyword>

2.2.1.2   <topcClas>   Topic Classification

- Optional
- Repeatable
- Attributes: <u>ID, xml:lang, source</u>, vocab, vocabURI

Description: The classification field indicates the broad substantive topic(s) that the data cover. Library of Congress subject terms may be used here. The "vocab" attribute is provided for specification of the controlled vocabulary in use, e.g., LCSH, MeSH, etc. The "vocabURI" attribute specifies the location for the full controlled vocabulary. Maps to Dublin Core Subject element. Inclusion of this element in the codebook is recommended.

Note:
        See Appendix C for more information on this tag.

<odesi>
format
Formatting Note 1:
        If possible, this list should be alphabetized.

<odesi>
format
Formatting Note 2:
        Only the first letter of the first word in a Topic Classifications phrase is capitalized, unless it is a proper name.

Example:
        <topcClas vocab="**Elections**</topcClas>
        <topcClas vocab="**Health**</topcClas>
        <topcClas vocab="**Public opinion**</topcClas>
        <topcClas vocab="**Religion**</topcClas>
        <topcClas vocab="**Smoking**</topcClas>

2.2.2   Abstract

- Optional
- Repeatable
- Attributes: <u>ID, xml:lang, source</u>, date

Description: An unformatted summary describing the purpose, nature, and scope of the data collection, special characteristics of its contents, major subject areas covered, and what questions the PIs attempted to answer when they conducted the study. A listing of major variables in the study is important here. In cases where a codebook contains more than one abstract (for example, one might be supplied by the data producer and another prepared by the data archive where the data are deposited), the source and date attributes may be used to distinguish the abstract versions. Maps to Dublin Core Description. Inclusion of this element is recommended. Date attribute should follow ISO convention of YYYY-MM-DD.

Note:
See Appendix C for more information on this tag.

Example:
<timePrd event="**start**" date="**1997-01-01**">**January 1, 1997**</timePrd>
<timePrd event="**end**" date="**1997-12-31**">**December 31, 1997**</timePrd>

2.2.3.2 <collDate> Date of Collection

- Optional
- Repeatable
- Attributes: ID, xml:lang, source, date, event, cycle

Description: Contains the date(s) when the data were collected. Use the event attribute to specify "start", "end", or "single" for each date entered to distinguish between, for example, the first day of collection (start), only day of collection (single), and last day of collection (end). The ISO standard for dates (YYYY-MM-DD) is recommended for use with the date attribute. The "cycle" attribute permits specification of the relevant cycle, wave, or round of data. Maps to Dublin Core Coverage. Inclusion of this element in the codebook is recommended.

Note:
See Appendix C for more information on this tag.

Example 1:
<collDate event="**single**" date="**1998-04-19**">**April 19, 1998**</collDate>

Example 2:
<collDate event="**start**" date="**1998-04-19**">**April 19, 1998**</collDate>
<collDate event="**end**" date="**1998-05-01**">**May 1, 1998**</collDate>

2.2.3.3 <nation> Country

- Optional
- Repeatable
- Attributes: ID, xml:lang, source, abbr

Description: Indicates the country or countries covered in the file. Attribute "abbr" may be used to list common abbreviations; use of ISO country codes is recommended. Maps to Dublin Core Coverage element. Inclusion of this element is recommended.

Example 1:
<nation>**Canada**</nation>

Example 2:
<nation abbr="**GB**">**Great Britain**</nation>

2.2.3.4   <geogCover>   Geographic Coverage

- Optional
- Repeatable
- Attributes: <u>ID, xml:lang, source</u>

Description: Information on the geographic coverage of the data. Include the total geographic scope of the data, and any additional levels of geographic coding provided in the variables. Maps to Dublin Core Coverage. Inclusion of this element is recommended.

Note:
　　See Appendix C for more information on this tag.

Example:
　　<geogCover>**Canada, Provinces and Territories**</geogCover>

2.2.3.5   <geogUnit>   Geographic Unit

- Optional
- Repeatable
- Attributes: <u>ID, xml:lang, source</u>

Description: Lowest level of geographic aggregation covered by the data.

Note:
　　See Appendix C for more information on this tag.

<odesi>
　format

Formatting Note:
　　No abbreviations should be used in this tag.

Example 1:
　　<geogUnit>**Province, Territory**</geogUnit>

Example 2:
　　<geogUnit>**Census Metropolitan Area**</geogUnit>

Example 3:
　　<geogUnit>**Health Region**</geogUnit>

2.2.3.8   <analyUnit>   Unit of Analysis

- Optional
- Repeatable
- Attributes: <u>ID, xml:lang, source</u>, unit

Description: Basic unit of analysis or observation that the file describes: individuals, families/households, groups, institutions/organizations, administrative units, etc. The "unit" attribute is included to permit the development of a controlled vocabulary for this element.

Note:
　　See Appendix C for more information on this tag.

Example 1:
    <anlyUnit>**Households**</anlyUnit>

Example 2:
    <anlyUnit>**Individual**</anlyUnit>

Example 3:
    <anlyUnit>**Episode**</anlyUnit>

Example 4:
    <anlyUnit>**Family**</anlyUnit>

2.2.3.9   <universe>   Universe

▪ Optional
▪ Repeatable
▪ Attributes: <u>ID, xml:lang, source</u>, level, clusion

Description: A description of the population covered by the data in the file; the group of persons or other elements that are the object of the study and to which the study results refer. Age, nationality, and residence commonly help to delineate a given universe, but any of a number of factors may be involved, such as age limits, sex, marital status, race, ethnic group, nationality, income, veteran status, criminal convictions, etc. The universe may consist of elements other than persons, such as housing units, court cases, deaths, countries, etc. In general, it should be possible to tell from the description of the universe whether a given individual or element (hypothetical or real) is a member of the population under study. Also known as universe of interest, population of interest, and target population. A "level" attribute is included to permit coding of the level to which universe applies, i.e., the study level, the file level (if different from study), or the variable level. The "clusion" attribute provides for specification of groups included (I) in or excluded (E) from the universe.

Note:
    See Appendix C for more information on this tag.

<odesi>
format

Formatting Note:
    If the words "Included" and "Excluded" are not at the beginning of the paragraphs, the user will not know what the paragraph means.

Example:
    <universe level="**study**" clusion="**I**"> **Included: The target population of the SHS is comprised of individuals living in private, Canadian households who are neither official representatives of foreign countries living in Canada and their families, nor residents of Indian reservations or Crown lands.**</universe>
    <universe level="**study**" clusion="**E**">**Excluded: The following groups were excluded from the survey; those living on Indian reserves and crown lands; official representatives of foreign countries living in Canada and their families; members of religious and other communal colonies; members of the Canadian Armed Forces living in Military Camps; people living in residences for senior citizens; people living full time in institutions: for example, inmates of penal institutions and chronic care patients living in hospitals and nursing homes. The survey covers about 98% of the population in the ten provinces. In the Territories, coverage was restricted to 88% in the Yukon, 92% in the Northwest Territories, and 89% in Nunavut. Information was not gathered from persons temporarily living away from their families (for example, students at university), because it would be gathered from their families if selected. In this way, double counting of such individuals was avoided**</universe>

2.2.3.10   <dataKind>   Kind of Data

- Optional
- Repeatable
- Attributes: <u>ID, xml:lang, source</u>

Description: The type of data included in the file: survey data, census/enumeration data, aggregate data, clinical data, event/transaction data, program source code, machine-readable text, administrative records data, experimental data, psychological test, textual data, coded textual, coded documents, time budget diaries, observation data/ratings, process-produced data, etc.

 Note:
    See Appendix C for more information on this tag.

<odesi>
format

Formatting Note:
    Only the first letter of the first word in the Type of Data phrase is capitalized.

Example 1:
    <dataKind>**Survey data**</dataKind >

Example 2:
    <dataKind>**Aggregate data**</dataKind >

Example 3:
    <dataKind>**Time budget diaries**</dataKind >

Example 4:
    <dataKind>**Census data**</dataKind >

Example 5:
    <dataKind>**Time series**</dataKind >

2.2.4   <notes>   Notes and Comments

- Optional
- Repeatable
- Attributes: <u>ID, xml:lang, source</u>, type, subject, level, resp, sdatrefs

Description: Used to indicate additional information regarding the scope of a data collection. The attributes for notes permit a controlled vocabulary to be developed (type and subject), the level of the DTD to which the note refers to be identified (study, file, variable, etc.), and the author of the note to be indicated (resp).

Note:
    See Appendix C for more information on this tag.

Example:
    <titl>**Survey of Consumer Finances, 1997: Individual File [Canada]**</titl>
    <notes>**The year listed is the year the survey was conducted, while the data refer to the previous year, known as the Reference Year. For this survey, it is 1996.**</notes>

2.3   <method>   Methodology

- Optional
- Repeatable
- Attributes: <u>ID, xml:lang, source</u>

Description: This section describes the methodology and processing involved in a data collection.

2.3.1   <dataColl>   Data Collection Methodology

- Optional
- Repeatable
- Attributes: <u>ID, xml:lang, source</u>

Description: Information about the methodology employed in a data collection.

2.3.1.1   <timeMeth>   Time Method

- Optional
- Repeatable
- Attributes: <u>ID, xml:lang, source</u>, method

Description: The time method or time dimension of the data collection. The "method" attribute is included to permit the development of a controlled vocabulary for this element.

Note:
    See Appendix C for more information on this tag.

Example 1:
    <timeMeth>**Cross-Sectional**</timeMeth>

Example 2:
    <timeMeth>**Longitudinal**</timeMeth>

Example 3:
    <timeMeth>**Pilot**</timeMeth>

Example 4:
    <timeMeth>**Time-series**</timeMeth>

Example 5:
    <timeMeth>**Trend study**</timeMeth>

Example 6:
    <timeMeth>**Panel survey**</timeMeth>

2.3.1.2   <dataCollector>   Data Collector

- Optional
- Repeatable
- Attributes: <u>ID, xml:lang, source</u>, abbr, affiliation

Description: The entity (individual, agency, or institution) responsible for administering the questionnaire or interview or compiling the data. This refers to the entity collecting the data, not to the entity producing the documentation. Attribute "abbr" may be used to list common abbreviations given to agencies, etc. Attribute "affiliation" may be used to record affiliation of the data collector.

Note 1:
    Unless you know the specific division, leave this tag blank. Only use it if you have explicit information.

Note 2:
    See Appendix C for more information on this tag.

<odesi>
    format

Formatting Note 1:
    Go from the larger entity to the smaller entity.

<odesi>
    format

Formatting Note 2:
    The note you have here should be repeated in every field that can take a corporate entity.

Example 1:
    <dataCollector abbr="**SRC**" affiliation="**University of Michigan**">**Survey Research Center**</dataCollector>

Example 2:
    <dataCollector abbr="" affiliation="**Statistics Canada**">**Special Surveys Division**</dataCollector>

Example 3:
    <dataCollector abbr="**CRIC**" affiliation="**Centre for Research on Canada**">**Environics Research Group**</dataCollector>

2.3.1.2   \<frequenc\>   Frequency of Data Collection

- Optional
- Repeatable
- Attributes: <u>ID, xml:lang, source</u>, freq

Description: If the data collected include more than one point in time, indicate the frequency with which the data were collected. The "frequency" attribute is included to permit the development of a controlled vocabulary for this element.

  Note 1:
    See Appendix C for more information on this tag.

  Note 2:
    Here are some of the different frequencies that could be used, followed by the definition:
- Quinquennial – every 5 years
- Biennial – every 2nd year
- Annual – every year
- Biannual – twice a year
- Monthly – every month
- Weekly – every week
- Occasional – conducted more than once, but not on a regular basis
- One-time – only conducted once

  Example:
    \<titl\>**Survey of Household Spending, 2003 [Canada]**\</titl\>
    \<frequenc\>**Annual 1997-1999**\</frequenc\>
    \<frequenc\>**Biennial 2000**\</frequenc\>

2.3.1.3   \<sampProc\>   Sampling Procedure

- Optional
- Repeatable
- Attributes: <u>ID, xml:lang, source</u>

Description: The type of sample and sample design used to select the survey respondents to represent the population. May include reference to the target sample size and the sampling fraction.

  Note:
    See Appendix C for more information on this tag.

\<odesi\>
   format

  Formatting Note:
    Only the first letter of the first word in the Sampling Procedure phrase is capitalized.

  Example 1:
    \<sampProc\>**Stratified random sample**\</sampProc\>

  Example 2:
    \<sampProc\>**Random digit dialling**\</sampProc\>

  Example 3:
    \<sampProc\>**Cluster sample**\</sampProc\>

Example 4:
<sampProc>**Two stage selection process: stage 1 is the selection of households; and stage 2 is selection of respondents within each household.**</sampProc>

2.3.1.6   <collMode>   Mode of Data Collection

- Optional
- Repeatable
- Attributes: <u>ID, xml:lang, source</u>

Description: The method used to collect the data; instrumentation characteristics.

Note:
See Appendix C for more information on this tag.

<odesi> format

Formatting Note:
Only the first letter of the first word in the Mode of Collection phrase is capitalized.

Example 1:
<collMode>**Face-to-face**</collMode>

Example 2:
<collMode>**Paper questionnaire**</collMode>

Example 3:
<collMode>**Computer assisted, random digit dialing, telephone interview**</collMode>

Example 4:
<collMode>**Web survey**</collMode>

2.3.1.7   <resInstru>   Type of Research Instrument

- Optional
- Repeatable
- Attributes: <u>ID, xml:lang, source</u>, type

Description: The type of data collection instrument used. "Structured" indicates an instrument in which all respondents are asked the same questions/tests, possibly with precoded answers. If a small portion of such a questionnaire includes open-ended questions, provide appropriate comments. "Semi-structured" indicates that the research instrument contains mainly open-ended questions. "Unstructured" indicates that in-depth interviews were conducted. The "type" attribute is included to permit the development of a controlled vocabulary for this element.

Example 1:
<resInstru>**structured**</resInstru>

Example 2:
<resInstru>**semi-structured**</resInstru>

Example 3:
<resInstru>**unstructured**</resInstru>

2.3.1.8   <sources>   Sources Statement

- Optional
- Not Repeatable
- Attributes: ID, xml:lang, source

Description: Description of sources used for the data collection. The element is nestable so that the sources statement might encompass a series of discrete source statements, each of which could contain the facts about an individual source. This element maps to Dublin Core Source element.

2.3.1.8.1   <dataSrc>   Data Source

- Optional
- Repeatable
- Attributes: ID, xml:lang, source

Description: Used to list the book(s), article(s), serial(s), and/or machine-readable data file(s)—if any—that served as the source(s) of the data collection.

Example 1:
   <titl>**Census of Canada, 1871**</titl>
   <dataSrc>**Canadian Historical Mobility Project, 1871, National Documentation for SPSS Portable File** </dataSrc>

Example 2:
   <titl>**The Trans-Atlantic Slave Trade: A Database on CD-ROM [In-House Use Only]**</titl>
   <dataSrc>**Sources for The Trans-Atlantic Slave Trade Database are many and varied. A list of sources is available in Appendix A of the User Guide.**</dataSrc>

Example 3:
   <titl>**Canadian historical mobility project**</titl>
   <altTitl>**Census of Canada, 1871**</altTitl>
   <docSrc>
    <titl>**Canadian historical mobility project, 1871, national documentation for SPSS portable file**</titl>
   </docSrc>
   <sources>
   <dataSrc>**The studies are based on samples taken from the nominal data given on the census manuscripts of 1861 and 1871.**</dataSrc>
   <arcOrig>
   At the time of writing census manuscripts of 1861 and 1871 are available, in varying quality, on microfilm from 1851 to 1881 for Canada, via the National Archives of Canada.</arcOrig>
   </sources>

Example 4:
   <dataSrc>**United States Internal Revenue Service Quarterly Payroll File**</dataSrc>

2.3.1.8.2   \<srcOrig\>   Origins of Sources

- Optional
- Repeatable
- Attributes: <u>ID, xml:lang, source</u>

Description: For historical materials, information about the origin(s) of the sources and the rules followed in establishing the sources should be specified. May not be relevant to survey data.

Example 1:
    \<titl\>**Census of Canada, 1871**\</titl\>
    \<srcOrig\>**The study is based on samples taken from the nominal data given on the census manuscripts of 1861 and 1871. Nominal data means simply the records of the individuals and households recorded on the original folios by the nineteenth century census enumerators.** \</srcOrig\>

Example 2:
    \<titl\>**The Trans-Atlantic Slave Trade: A Database on CD-ROM [In-House Use Only]**\</titl\>
    \<srcOrig\>**The study draw on a wide range of published and archival information from Brazil, Denmark, Great Britain, Portugal, and United States.** \</srcOrig\>

2.3.1.9   \<collSitu\>   Characteristics of the Data Collection Situation

- Optional
- Repeatable
- Attributes: <u>ID, xml:lang, source</u>

Description: Description of **noteworthy** aspects of the data collection situation. Includes information on factors such as cooperativeness of respondents, duration of interviews, number of call-backs, etc.

Note:
    See Appendix C for more information on this tag.

Example:
    \<collSitu\>**There were 1,194 respondents who answered questions in face-to-face interviews lasting approximately 75 minutes each.**\</collSitu\>

2.3.1.12   \<weight\>   Weighting

- Optional
- Repeatable
- Attributes: <u>ID, xml:lang, source</u>

Description: The use of sampling procedures may make it necessary to apply weights to produce accurate statistical results. Describe here the criteria for using weights in analysis of a collection.

Note 1:
    See Appendix C for more information on this tag.

Note 2:
> If there is a brief description of the weight, use it here.

Note 3:
> If no weight variable is included but the User's Guide tells how to calculate it, include this information.

Note 4:
> Do not include any formulas.

Example 1:
> <weight>**WEIGHT=Weight at household level** </weight>

Example 2:
> <titl>**General Social Survey, 1998 [Canada]: Cycle 12, Time Use, Main File**</titl>
> <weight>**WGHTFIN=the number of persons in the population that a record on the main file represents**</weight>

Example 3:
> <titl> **Household Internet Use Survey, 2003 [Canada]**</titl>
> <weight>**The principles behind the calculation of the weights for the HIUS are nearly identical to those for the LFS. However, this survey is a household-weighted survey, not a person-weighted survey. Also, further adjustments are made to the LFS sub-weights in order to derive a final weight for each record on the HIUS microdata file. 1) An adjustment to account for the use of a four-sixths sub-sample, instead of the full LFS sample. 2) An adjustment to account for the additional non-response to the supplementary survey, i.e., households that did not respond to the HIUS but did respond to the LFS (or for which previous month's LFS data was brought forward). Statistical techniques are used to group together records that are similar in terms of demographic variables obtained from LFS responses. The adjustment is made separately within all non-response groups created for each province. Household Internet Use Survey, 2003 – user guide 3) The final adjustment ensured that estimates produced for a province-household size group would agree with the known population totals for that province-household size group. The adjustments were made for household size groupings of one person, two people and three or more people. The resulting weight (WTHM) is the final weight which appears on the HIUS microdata file.** </weight>

Example 4:
> <titl>**General Social Survey, 1998 [Canada]: Cycle 12, Time Use, Episode File**</titl>
> <weight>**WGHTEPI=the number of time use episodes that a record on the episode file represents**</weight>

2.3.2   <notes>   Notes and Comments

- Optional
- Repeatable
- Attributes: <u>ID, xml:lang, source</u>, type, subject, level, resp, sdatrefs

Description: For clarifying information/annotation regarding the parent element. The attributes for notes permit a controlled vocabulary to be developed ("type" and "subject"), indicate the "level" of the DDI to which the note applies (study, file, variable, etc.), and identify the author of the note ("resp").

Example:
　　<notes>**Additional information on derived variables has been added to this marked-up version of the documentation**.</notes>

2.3.3　　　　　　<anlyInfo>　　　　　　Data Appraisal

- Optional
- Not Repeatable
- Attributes: ID, xml:lang, source

Description: Information on data appraisal.

2.3.3.1　　　　　　<respRate>　　　　　　Response Rate

- Optional
- Repeatable
- Attributes: ID, xml:lang, source

Description: The percentage of sample members who provided information, if available.

Note:
　　See Appendix C for more information on this tag.

Example 1:
　　<respRate> **For the 1996 Sun Exposure Survey, 12,065 telephone numbers were called. 5,847 of these were assumed to belong to households and 6,218 belonged to non-households. Of the 5,847 numbers, 231 were called repeatedly, but no contact was made; for the purpose of weighting and calculating response rates, we assume that these numbers belong to households. 1,589 of the households were non-responding because either they refused or the selected respondent could not be reached during the survey collection period. A complete interview was administered to the selected respondent in the remaining 4,027 households. However, due to incomplete roster information (age/sex data), 4 respondents were dropped from the sample. The final sample, of respondents 15 years of age and over, totaled 4,023.**</respRate>

Example 2:
　　<respRate>**For 1993, the estimated inclusion rate for TEDS-eligible providers was 91 percent, with the inclusion rate for all treatment providers estimated at 76 percent (including privately and publicly funded providers).**</respRate>

2.3.3.2　　　　　　<EstSmpErr>　　　　　　Estimates of Sampling Error

- Optional
- Repeatable
- Attributes: ID, xml:lang, source

Description: Measure of how precisely one can estimate a population value from a given sample.

Example:

<EstSmpErr>**Since it is an unavoidable fact that estimates from a sample survey are subject to sampling error, sound statistical practice calls for researchers to provide users with some indication of the magnitude of this sampling error. The basis for measuring the potential size of sampling errors is the standard deviation of the estimates derived from survey results. However, because of the large variety of estimates that can be produced from a survey, the standard deviation of an estimate is usually expressed relative to the estimate to which it pertains. This resulting measure, known as the coefficient of variation (CV) of an estimate, is obtained by dividing the standard deviation of the estimate by the estimate itself and is expressed as a percentage of the estimate.**

Statistics Canada commonly uses CV results when analyzing data and urges users producing estimates from the CCHS Cycle 3.1 data files to also do so. For details on how to determine CVs, see Section 11. For guidelines on how to interpret CV results, see the table at the end of Subsection 10.4.
http://nesstar.tdr.uoguelph.ca/HEALTH/CCHS/2005/DOCS/guide_e.pdf</EstSmpErr>

Explanation of Example:
> This example is from the CCHS 2005. The information is taken directly from the User Guide.

2.4   <dataAccs>   Access to Data

- Optional
- Repeatable
- Attributes: <u>ID, xml:lang, source</u>

Description: This section describes access conditions and terms of use for the data collection. In cases where access conditions differ across individual files or variables, multiple access conditions can be specified. The access conditions applying to a study, file, variable group, or variable can be indicated by an IDREF attribute on the study (2.0), file (3.0), variable group (4.1), or variable (4.2) elements called "access".

2.4.1   <setAvail>   Data Collection Availability

- Optional
- Repeatable
- Attributes: <u>ID, xml:lang, source</u>, media, callno, label, type

Description: Information on availability and storage of the collection. The "media" attribute may be used in combination with any of the sub-elements. See Location of Data Collection.

 Example:
<setAvail media="**On-line subscription service**">

2.4.1.1   <accsPlac>   Location of Data Collection

- Optional
- Repeatable
- Attributes: <u>ID, xml:lang, source</u>, URI

Description: Location where the data collection is currently stored. Use the URI attribute to provide a URN or URL for the storage site or the actual address from which the data may be downloaded.

Note:
Use this tag to direct users to either a download site for a database or to come into your work place for access.

Example 1:
<accsPlac URI="**http://nesstar.tdr.uoguelph.ca/AGRIC/RT/DOCS/rt-tables.html/**">
Access Excel Tables Here </accsPlac>

Example 2:
<accsPlac URI="**http://www.biblio.uottawa.ca.proxy.library.carleton.ca/content-page.php?g=en&s=gsg&c=src-justice**/">**Access the Canadian Centre for Justice Statistics Here**</accsPlac>

Example 3:
<accsPlac URI="**http://new.sourceoecd.org/**">**Access SourceOECD Subscription Service Here**</accsPlac>

2.4.1.2   <origArch>   Archive Where Study Originally Stored

- Optional
- Not Repeatable
- Attributes: ID, xml:lang, source

Description: Archive from which the data collection was obtained; the originating archive.

Example:
<titl>**Canadian Gallup Poll, May 1949, #186**</titl>
<origArch>**Carleton University. Data Centre**</origArch>

2.4.1.3   <avlStatus>   Availability Status

- Optional
- Repeatable
- Attributes: ID, xml:lang, source

Description: Statement of collection availability. An archive may need to indicate that a collection is unavailable because it is embargoed for a period of time, because it has been superseded, because a new edition is imminent, etc. It is anticipated that a controlled vocabulary will be developed for this element.

Example:
<avlStatus>**This collection is superseded by CENSUS OF POPULATION, 1880 [UNITED STATES]: PUBLIC USE SAMPLE (ICPSR 6460).**</avlStatus>

2.4.1.4   \<collSize\>   Extent of Collection

- Optional
- Not Repeatable
- Attributes: <u>ID, xml:lang, source</u>

Description: Summarizes the number of physical files that exist in a collection, recording the number of files that contain data and noting whether the collection contains machine-readable documentation and/or other supplementary files and information such as data dictionaries, data definition statements, or data collection instruments.

Example:
\<collSize\>**1 data file + machine-readable documentation (PDF) + SAS data definition statements**\</collSize\>

2.4.1.5   \<complete\>   Completeness of Study Stored

- Optional
- Not Repeatable
- Attributes: <u>ID, xml:lang, source</u>

Description: This item indicates the relationship of the data collected to the amount of data coded and stored in the data collection. Information as to why certain items of collected information were not included in the data file stored by the archive should be provided.

Example:
\<complete\>**Because of embargo provisions, data values for some variables have been masked. Users should consult the data definition statements to see which variables are under embargo. A new version of the collection will be released by ICPSR after embargoes are lifted.**\</complete\>

2.4.1.6   \<fileQnty\>   Number of Files

- Optional
- Not Repeatable
- Attributes: <u>ID, xml:lang, source</u>

Description: Total number of physical files associated with a collection.

Example:
\<fileQnty\> **5 files**\</fileQnty\>

2.4.1.7   \<notes\>   Notes and Comments

- Optional
- Repeatable
- Attributes: <u>ID, xml:lang, source</u>, type, subject, level, resp, sdatrefs

Description: For clarifying information/annotation regarding the parent element. The attributes for notes permit a controlled vocabulary to be developed (type and subject), the level of the DTD to which the note refers to be identified (study, file, variable, etc.), and the author of the note to be indicated (resp).

2.4.2   <useStmt>   Use Statement

- Optional
- Repeatable
- Attributes: ID, xml:lang, source

Description: Information on terms of use for the data collection,

2.4.2.3   <restrctn>   Restrictions of Use

- Optional
- Not Repeatable
- Attributes: ID, xml:lang, source

Description: Any restrictions on access to or use of the collection such as privacy certification or distribution restrictions should be indicated here. These can be restrictions applied by the author, producer, or disseminator of the data collection. If the data are restricted to only a certain class of user, specify which type.

Note:
An external link may be added for more detailed restrictions or information by the institution.

Example 1:
<restrctn>**The data may be used by current students, faculty and staff of the University of Guelph, the University of Waterloo and Wilfrid Laurier University for academic and personal research only**.</restrctn>

Example 2:
<restrctn>**For members of the Data Liberation Initiative community. For academic use only, that is, for teaching and non-commercial research. Data may not be re-disseminated**. </restrctn>

Example 3:
<restrctn> **<As a rule, agreements stipulate that the data may be used by students, faculty and staff of the University of Guelph, the University of Waterloo and Wilfrid Laurier University for academic and personal research only. Any use of these data for commercial applications or contract purposes will violate our licences. Use by members of the community outside of these three institutions could also violate our licences. Please contact staff at any of the three Centres for more information or visit>**
<ExtLink URI="http://tdr.uoguelph.ca/tdrlic.htm" title="TDR Licensing Agreements"/> </restrctn>

Example 4:
<restrctn>**This data is restricted to use by Queen's University faculty, staff and students for academic research only. You must register to use this data.</restrctn>**
<ExtLink URI=http://db.library.queensu.ca/register/datareq2.htm title="Registration Page" /></restrctn>

2.4.2.4   \<contact\>   Contact Persons

- Optional
- Repeatable
- Attributes: <u>ID, xml:lang, source</u>, affiliation, URI, email

Description: Name of institution(s) responsible for the work and their address. The URI attribute should be used to indicate a URN or URL for the homepage of the contact institution. The email attribute is used to indicate an email address for the contact institution.

Example 1:
  \<contact affiliation="**University of Guelph**" email="**drc@listserv.uoguelph.ca**">
  \</contact\>

Example 2:
  \<contact affiliation="**Carleton University**" email="**ssdata@ccs.carleton.ca**">\</contact\>

Example 3:
  \<contact affil="**Statistics Canada**" URI=" **http://www.statcan.ca/english/Dli/dli.htm**>
  email="**dli-idd@statcan.ca**">**Data Liberation Initiative**\</contact\>

Example 4:
  \<contact affil="**Statistics Canada**" URI=**http://www.statcan.ca>**
  email=ssd@statcan.ca">**Client Services Section. Special Surveys**
  Division\</contact\>

2.4.2.5   \<citReq\>   Citation Requirement

- Optional
- Not Repeatable
- Attributes: <u>ID, xml:lang, source</u>

Description: Text of requirement that a data collection should be cited properly in articles or other publications that are based on analysis of the data.

Example 1: (taken from the DLI Licensing Agreement, Section 9)
  \<citReq>**The publishing of analysis and results from research using any of the data products is permitted in research communications such as scholarly papers, journals and the like. The authors of these communications are required to cite Statistics Canada as the source of the data, and to indicate that the results or views expressed are those of the author/ authorized user and are not those of Statistics Canada. Permission to include extracts of these data in textbooks must be obtained from the Licencing Section of Statistics Canada's Marketing Division.**\</citReq\>

Example 2:
  \<titl\> **Canadian Gallup Poll, May 1949, #186** \</titl\>
  \<citReq>**The publishing of analysis and results from research using this data is permitted in research communications such as scholarly papers, journals and the like. The authors of these communications are required to cite Gallup Canada Inc. as the source of the data, and to indicate that the results or views expressed are those of the author/ authorized user and are not those of Gallup Canada Inc.**
  \</citReq\>

Example 3:
<citReq>**Publications based on ICPSR data collections should acknowledge those sources by means of bibliographic citations. To ensure that such source attributions are captured for social science bibliographic utilities, citations must appear in footnotes or in the reference section of publications**.</citReq>

2.4.2.7   <conditions>   Conditions of Use

- Optional
- Not Repeatable
- Attributes: <u>ID, xml:lang, source</u>

Description: Indicate any additional information that will assist the user in understanding the access and use conditions of the data collection.

Example:
<conditions>**Data Liberation Initiative** <ExtLink URI="**http://tdr.tug-libraries.on.ca/HELPS/dlilicence.html**" title="**DLI Licence Agreement**" /></conditions>

2.4.2.8   <disclaimer>   Disclaimer

- Optional
- Not Repeatable
- Attributes: <u>ID, xml:lang, source</u>

Description: Information regarding responsibility for uses of the data collection.

Example 1:
<titl>**Canadian Gallup Poll, May 1949, #186** </titl>
<disclaimer> **The original collector of the data, Gallup Canada Inc., bears no responsibility for uses of this collection or for interpretations or inferences upon such uses.** </disclaimer>

Example 2:
<disclaimer>**The original collector of the data, ICPSR, and the relevant funding agency bear no responsibility for uses of this collection or for interpretations or inferences based upon such uses.**</disclaimer>

2.5   <othrStdyMat>   Other Study Description Materials

- Optional
- Repeatable
- Attributes: <u>ID, xml:lang, source</u>

Description: Other materials relating to the study description. This section describes other materials that are related to the study description that are primarily descriptions of the content and use of the study, such as appendices, sampling information, weighting details, methodological and technical details, publications based upon the study content, related studies or collections of studies, etc. This section may point to other materials related to the description of the study through use of the generic citation element (A.6), which is available for each element in this section. <othrStdyMat> maps to Dublin Core Relation element.

Note that Section 5.0, Other Study-Related Materials, should be used for materials used in the production of the study or useful in the analysis of the study. The materials in Section 5.0 may be entered as PCDATA (ASCII text) directly into the document (through use of the txt element). That section may also serve as a "container" for other electronic materials by providing a brief description of the study-related materials accompanied by the "type" and "level" attributes further defining the materials. Other Study-Related Materials in Section 5.0 may include: questionnaires, coding notes, SPSS/SAS/Stata setup files (and others), user manuals, continuity guides, sample computer software programs, glossaries of terms, interviewer/project instructions, maps, database schema, data dictionaries, show cards, coding information, interview schedules, missing values information, frequency files, variable maps, etc.

2.5.1   <relMat>   Related Materials

- Optional
- Repeatable
- Attributes: <u>ID, xml:lang, source</u>, callno, label, media, type

Description: Describes materials related to the study description, such as appendices, additional information on sampling found in other documents, etc. Can take the form of bibliographic citations. This element can contain either PCDATA or a citation or both, and there can be multiple occurrences of both the citation and PCDATA within a single element. May consist of a single URI or a series of URIs comprising a series of citations/references to external materials which can be objects as a whole (journal articles) or parts of objects (chapters or appendices in articles or documents).

Example:
<relMat> **Full details on the research design and procedures, sampling methodology, content areas, and questionnaire design, as well as percentage distributions by respondent's sex, race, region, college plans, and drug use, appear in the annual ISR volumes MONITORING THE FUTURE: QUESTIONNAIRE RESPONSES FROM THE NATION'S HIGH SCHOOL SENIORS**.</relMat>

<relMat>**Current Population Survey, March 1999: Technical Documentation includes an abstract, pertinent information about the file, a glossary, code lists, and a data dictionary. One copy accompanies each file order. When ordered separately, it is available from Marketing Services Office, Customer Service Center, Bureau of the Census, Washington, D.C. 20233**. </relMat>

<relMat>**A more precise explanation regarding the CPS sample design is provided in Technical Paper 40, The Current Population Survey: Design and Methodology. Chapter 5 of this paper provides documentation on the weighting procedures for the CPS both with and without supplement questions**.</relMat>

2.5.2   \<relStdy>   Related Studies

- Optional
- Repeatable
- Attributes: ID, xml:lang, source

Description: Information on the relationship of the current data collection to others (e.g., predecessors, successors, other waves or rounds) or to other editions of the same file. This would include the names of additional data collections generated from the same data collection vehicle plus other collections directed at the same general topic. Can take the form of bibliographic citations.

 Example 1:
   \<titl>**General Social Survey, 2005 [Canada]: Cycle 19, Time Use, Main File**\</titl>
   \<relStdy>**For the General Social Survey, 2005, Cycle 19, Time Use, there are three related studies titled: General Social Survey, 1998, Cycle 12, Time Use: [Canada]; General Social Survey, 1992, Cycle 7, Time Use; [Canada]; and General Social Survey, 1986, Cycle 2, Time Use: [Canada]**.\</relStdy>

 Example 2:
   \<relStdy>**ICPSR distributes a companion study to this collection titled FEMALE LABOR FORCE PARTICIPATION AND MARITAL INSTABILITY, 1980: [UNITED STATES] (ICPSR 9199).**\</relStdy>

2.5.3   \<relPubl>   Related Publications

- Optional
- Repeatable
- Attributes: ID, xml:lang, source

Description: Bibliographic and access information about articles and reports based on the data in this collection. Can take the form of bibliographic citations.

 Example:
   \<relPubl>**Economic Behavior Program Staff. SURVEYS OF CONSUMER FINANCES. Annual volumes 1960 through 1970. Ann Arbor, MI: Institute for Social Research**.\</relPubl>

   \<relPubl>**Data from the March Current Population Survey are published most frequently in the Current Population Reports P- 20 and P- 60 series. These reports are available from the Superintendent of Documents, U. S. Government Printing Office, Washington, DC 20402. They also are available on the INTERNET at http:// www. census. gov. Forthcoming reports will be cited in Census and You, the Monthly Product Announcement (MPA), and the Bureau of the Census Catalog and Guide**. \</relPubl>

SECTION 3.0   DATA FILES DESCRIPTION
<fileDscr>

Tags in this section relate directly to the format and content of the datafiles. Note that the tags should correspond to the format you are working with – raw datafile, SAS dataset, NSDstat file.

"The File Description consists of information about the particular data file(s) containing numeric and/or numeric + textual information that the DDI-compliant file describes. This section consists of items describing the characteristics and contents of file(s) that comprise the study as described in the Study Description. There may be multiple file descriptions if there are multiple files in the collection."

Tags and corresponding DTD numbers are from DDI Version 2.0 codebook.

Source: DDI Codebook
http://www.icpsr.umich.edu/DDI/dtd/version2-1-all.html
http://www.icpsr.umich.edu/DDI/dtd/version2-1-tree.html

DTD Numbers	Tags
3.0	<fileDscr>
3.1	<fileTxt>
3.1.1	<fileName>
3.1.2	<fileCont>
3.1.3	<fileStrc>
3.1.3.1	<recGrp >
3.1.4	<dimensns>
3.1.4.1	<caseQnty>
3.1.4.2	<varQnty>
3.1.4.3	<logRecL>
3.1.4.5	<recNumTot>
3.1.5	<fileType>
3.1.6	<format>
3.1.8	<dataChck>
3.1.12	<verStmt>
3.1.12.1	<version>
3.1.12.2	<verResp>
3.1.12.3	<notes >
3.3	<notes >

Description of Tags and Working Examples

3.0   <fileDscr>   Data File Description

- Optional
- Repeatable
- Attributes: <u>ID, xml:lang, source</u>, URI, sdatrefs, methrefs, pubrefs, access

Description: Information about the data file(s) that comprises a collection. This section can be repeated for collections with multiple files. The "URI" attribute may be a URN or a URL that can be used to retrieve the file. The "sdatrefs" are summary data description references that record the ID values of all elements within the summary data description section of the Study Description that might apply to the file. These elements include: time period covered, date of collection, nation or country, geographic coverage, geographic unit, unit of analysis, universe, and kind of data. The "methrefs" are methodology and processing references that record the ID values of all elements within the study methodology and processing section of the Study Description that might apply to the file. These elements include information on data collection and data appraisal (e.g., sampling, sources, weighting, data cleaning, response rates, and sampling error estimates). The "pubrefs" attribute provides a link to publication/citation references and records the ID values of all citations elements within Other Study Description Materials or Other Study-Related Materials that pertain to this file. "Access" records the ID values of all elements in the Data Access section that describe access conditions for this file.

Remarks: When a codebook documents two different physical instantiations of a data file, e.g., logical record length (or OSIRIS) and card-image version, the Data File Description should be repeated to describe the two separate files. An ID should be assigned to each file so that in the Variable section the location of each variable on the two files can be distinguished using the unique file IDs

3.1   <fileTxt>   File-by-File Description

- Optional
- Repeatable
- Attributes: <u>ID, xml:lang, source</u>

Description: Information about the data file. Demonstrates that the next section of the codebook will deal with the datafile.

3.1.1   <fileName>   File Name

- Optional
- Not Repeatable
- Attributes: <u>ID, xml:lang, source</u>

Description: Contains a short title that will be used to distinguish a particular file/part from other files/parts in the data collection.

Note 1:
    This is the title of the datafile, which is not necessarily the same as the title of the survey.

Example 1:
   <fileName>**CTUMS 2006 C1 Person File**</fileName>

Example 2:
   <fileName> **GSS 2001: Cycle 15, Family History, Child File** </filename>

Example 3:
   <fileName>**Sun Exposure Survey 1996**</fileName>

3.1.2   <fileCont>   Contents of Files

- Optional
- Not Repeatable
- Attributes: <u>ID, xml:lang, source</u>

Description: Abstract or description of the datafile. A summary describing the purpose, nature, and scope of the data file, special characteristics of its contents, major subject areas covered, and what questions the PIs attempted to answer when they created the file. A listing of major variables in the file is important here. In the case of multi-file collections, this uniquely describes the contents of each file.

Example:
   <fileCont>**The annual HIUS contains detailed data on the Internet activities of Canadian household.** </fileCont>

3.1.3   <fileStrc>   File Structure

- Optional
- Not Repeatable
- Attributes: <u>ID, xml:lang, source</u>, type

Description: Type of file structure. Use attribute of "type" to indicate hierarchical, rectangular, or relational (the default is rectangular).

Note:
   See Appendix B for more information on this tag.

Example 1:
   <fileStrc>**Hierarchical**</fileStrc>

Example 2:
   <fileStrc>**Rectangular**</fileStrc>

3.1.3.1   <recGrp >   Record or Record Group

- Optional
- Repeatable
- Attributes: <u>ID, xml:lang, source</u>, recGrp, rectype, keyvar, rtypeloc, rtypewidth, rtypevtype, recidvar

Description: Used to describe record groupings if the file is hierarchical or relational. The attribute "recGrp" allows a record group to indicate subsidiary record groups that nest underneath; this allows for the encoding of a hierarchical structure of record groups. The attribute "rectype" indicates the type of record.

Example:
   <recGrp>**Person-level records**</recGrp>

3.1.4   <dimensns>   File Dimensions

- Optional
- Not Repeatable
- Attributes: <u>ID, xml:lang, source</u>

Description: Dimensions of the overall file.

3.1.4.1   <caseQnty>   Number of cases / Record Quantity

- Optional
- Repeatable
- Attributes: <u>ID, xml:lang, source</u>

Description: Number of cases or observations in the entire file. This is to be used for rectangular files only.

Note:
   See Appendix B for more information on this tag.

Example:
    <titl>**Survey of Household Spending, 2001, 1996 [Canada]**</titl>
    <caseQnty>**16901**</caseQnty>

3.1.4.2   <varQnty>   Number of variables per record

- Optional
- Repeatable
- Attributes: <u>ID, xml:lang, source</u>

Description: Number of variables in the entire file. This is to be used for rectangular files only.

Note:
   See Appendix B for more information on this tag.

Example:
    <titl>**Survey of Household Spending, 2001, 1996 [Canada]**</titl>
    <varQnty>**255**</varQnty>

3.1.4.3   <logRecL>   Record Length / Logical Record Length

- Optional
- Repeatable
- Attributes: <u>ID, xml:lang, source</u>

Description: Logical record length of the file, i.e., number of characters of data in the record. Only to be used for rectangular files or if all records in a hierarchical file are the same length.

Note:
    See Appendix B for more information on this tag.

Example:
    <titl>**Survey of Household Spending, 2001, 1996 [Canada]**</titl>
    <logRecL>**2093**</logRecL>

3.1.4.5   <recNumTot>   Overall Number of Records

- Optional
- Repeatable
- Attributes: <u>ID, xml:lang, source</u>

Description: Overall record count in the file. Particularly helpful in instances such as files with multiple cards/decks or records per case.

Note:
    See Appendix B for more information on this tag.

Example:
    <titl>**Survey of Household Spending, 2001 [Canada]**</titl>
    <recNumTot>**16,901**</recNumTot>

3.1.5   <fileType>   Type of File

- Optional
- Not Repeatable
- Attributes: <u>ID, xml:lang, source</u>, charset

Description: Types of data files include raw data (ASCII, EBCDIC, etc.) and software-dependent files such as SAS datasets, SPSS export files, etc. If the data are of mixed types (e.g., ASCII and packed decimal), state that here. Note that the element varFormat permits specification of the data format at the variable level. The "charset" attribute allows one to specify the character set used in the file, e.g., US-ASCII, EBCDIC, UNICODE UTF-8, etc.

Note:
    See Appendix B for more information on this tag.

Example 1:
    <titl>**Survey of Household Spending, 2001, 1996 [Canada]**</titl>
    <fileType>**shs2001.sav**</fileType>   *for an SPSS-Windows datafile

Example 2:
　　<titl>**Survey of Household Spending, 2001, 1996 [Canada]**</titl>
　　<fileType>**shs2001.sas7dbat**</fileType>　　　　*for a SAS dataset*

Example 3:
　　<titl>**Survey of Household Spending, 2001, 1996 [Canada]**</titl>
　　<fileType>**PUMDFSHS2001.txt**</fileType>　　　　*for a raw datafile*

3.1.6　　　　　　<format>　　　　　　Data Format

- Optional
- Not Repeatable
- Attributes: <u>ID, xml:lang, source</u>

Description: Physical format of the data file: Logical record length format, card-image format (i.e., data with multiple records per case), delimited format, free format, etc.

Note:
　See Appendix B for more information on this tag.

Example:
　<format>**comma-delimited**</format>

3.1.8　　　　<dataChck>　　　　　Extent of Processing Checks

- Optional
- Repeatable
- Attributes: <u>ID, xml:lang, source</u>

Description: Indicate here, at the file level, the types of checks and operations performed on the data file. A controlled vocabulary may be developed for this element in the future. The following examples are based on ICPSR's Extent of Processing scheme:

Example 1:
　　<dataChck>**The archive produced a codebook for this collection**.</dataChck>

Example 2:
　<dataChck>**Consistency checks were performed by Data Producer/ Principal Investigator**. </dataChck>

3.1.12　　　　<verStmt>　　　　　　Version Statement

- Optional
- Not Repeatable
- Attributes: <u>ID, xml:lang, source</u>

Description: Version statement for the data file, if one of a multi-file collection.

3.1.12.1   <version>   Version

- Optional
- Not Repeatable
- Attributes: <u>ID, xml:lang, source</u>, date, type

Description: Also known as release or edition. If there have been substantive changes in the file since its creation, this statement should be used. The ISO standard for dates (YYYY-MM-DD) is recommended for use with the date attribute.

<odesi>
format

Formatting Note:
    No abbreviations of acronyms should be used here.

Example:
    <version type='**revision**' date=**'2004-02-05'**>**Second Revision of Survey of Household Spending data**</version>

3.1.12.2   <verResp>   Version Responsibility Statement

- Optional
- Not Repeatable
- Attributes: <u>ID, xml:lang, source</u>, affiliation

Description: Used to indicate the organization or person responsible for the version of the file.

Example:
    <verResp> **Statistics Canada. Income Statistics Division**</verResp>

3.1.12.3   <notes >   Notes and Comments

- Optional
- Repeatable
- Attributes: <u>ID, xml:lang, source</u>, type, subject, level, resp, sdatrefs

Description: Used to indicate additional information regarding the version or version responsibility statement, in particular to indicate what makes a new version different from its predecessor. The attributes for notes permit a controlled vocabulary to be developed ("type" and "subject"), indicate the "level" of the DDI to which the note applies (study, file, variable, …), and identify the author of the note ("resp").

Note:
    Use this section when the variable labels, value labels and/or the missing values are changed.

Example:
    <notes>**Variable labels and value labels were edited.**</notes>

3.3   <notes>   Notes and Comments

- Optional
- Repeatable
- Attributes: <u>ID, xml:lang, source</u>, type, subject, level, resp, sdatrefs

Description: For clarifying information/annotation regarding the parent element. The attributes for notes permit a controlled vocabulary to be developed ("type" and "subject"), indicate the "level" of the DDI to which the note applies (study, file, variable, …), and identify the author of the note ("resp").

Note:
   This section will contain any notes about the data file not mentioned elsewhere in this document.

Example 1:
   <notes>**Data are collected quarterly, but only an annual file is distributed**. </notes>

SECTION 4.0   VARIABLE DESCRIPTION   \<dataDscr\>

Tags in this section relate to the variables and any statistic calculated from the data.

 "The Variable Description consists of a section describing variable groups and a section describing individual variables within the data file. The variable groups are defined as variables that may share common subjects, arise from the interpretation of a single question, or are linked by some other factor. The Variable Description is an extremely rich set of elements allowing for detailed descriptive information about response and analysis units, question text, forward progression and back flow, interviewer instructions, universe, valid and invalid data ranges, derived variables, summary statistics, etc. References to other parts of the DDI-compliant documentation file are possible through the use of IDREFS and links so that interrelationships among the elements may be used and documented."

Tags and corresponding DTD numbers are from DDI Version 2.0 codebook.

Note 1: The tag numbers may be different in this section depending on the version of DDI which you are using. If this is the case, merely look at the definition of the tag and go from there.

Note 2: Follow the variable groups that are used in the documentation. If there are no groups listed, look at the questionnaire. Usually the variables are grouped in it according to subject so these groupings can be used for this document.

Source: DDI Codebook
http://www.icpsr.umich.edu/DDI/dtd/version2-1-all.html
http://www.icpsr.umich.edu/DDI/dtd/version2-1-tree.html

DTD Numbers	Tags
4.0	\<dataDscr\>
4.1	\<varGrp\>
4.1.1	\<labl \>
4.1.6	\<notes\>
4.3	\<var\>
4.3.1	\<location\>
4.3.2	\<labl\>
4.3.8	\<qstn\>
4.3.8.1	\<preQTxt\>
4.3.8.2	\<qstnLit\>
4.3.8.3	\<postQTxt\>
4.3.8.4	\<forward\>
4.3.8.5	\<backward\>
4.3.8.6	\<ivuInstr\>
4.3.9	\<valrng\>
4.3.9.2	\<notes\>
4.3.10	\<invalrng\>
4.3.10.2	\<notes\>
4.3.11	\<undocCod\>
4.3.12	\<universe\>
4.3.13	\<TotlResp\>
4.3.14	\<sumStat\>
4.3.16	\<stdCatgry\>
4.3.18	\<catgry\>

4.3.18.1	<catValu>
4.3.18.2	<labl>
4.3.18.4	<catStat>
4.3.19	<codInstr>
4.3.22	<derivation>
4.3.22.1	<drvdesc>
4.3.22.2	<drvcmd>
4.3.23	<varFormat>
4.3.26	<notes>
4.5	<notes>

Description of Tags and Working Examples

4.0   <dataDscr>   Variable Description

- Optional
- Repeatable
- Attributes: <u>ID, xml:lang, source</u>

Description: Shows that the section dealing with the data is beginning.

4.1   <varGrp>   Variable Group

- Optional
- Repeatable
- Attributes: <u>ID, xml:lang, source</u>, type, var, varGrp, name, sdatrefs, methrefs, pubrefs, access, nCube

Description: A group of variables that may share a common subject, arise from the interpretation of a single question, or are linked by some other factor.

Attributes within the <varGrp>
- ➤ <u>ID</u>
  To uniquely identify the Variable Group

- ➤ <u>Type</u>
  The "type" of group attribute refers to the general type of grouping of the variables, e.g., subject, multiple response.
  Specific variable groups, included within the 'type' attribute, are:
  - **Section:** Questions which derive from the same section of the questionnaire, e.g., all variables located in Section C.
  - **Multiple response:** Questions where the respondent has the opportunity to select more than one answer from a variety of choices, e.g., what newspapers have you read in the past month (with the respondent able to select up to five choices).
  - **Grid:** Sub-questions of an introductory or main question but which do not constitute a multiple response group, e.g., I am going to read you some events in the news lately and you tell me for each one whether you are very interested in the event, fairly interested in the fact, or not interested in the event.
  - **Display:** Questions which appear on the same interview screen (CAI) together or are presented to the interviewer or respondent as a group.
  - **Repetition:** The same variable (or group of variables) which are repeated for different groups of respondents or for the same respondent at a different time.
  - **Subject:** Questions which address a common topic or subject, e.g., income, poverty, children.
  - **Version:** Variables, often appearing in pairs, which represent different aspects of the same question, e.g., pairs of variables (or groups) which are adjusted/unadjusted for inflation or season or whatever, pairs of variables with/without missing data imputed, and versions of the same basic question.
  - **Iteration:** Questions that appear in different sections of the data file measuring a common subject in different ways, e.g., a set of variables which report the progression of respondent income over the life course.

- **Analysis:** Variables combined into the same index, e.g., the components of a calculation, such as the numerator and the denominator of an economic statistic.
- **Pragmatic**: A variable group without shared properties.
- **Record:** Variable from a single record in a hierarchical file.
- **File:** Variable from a single file in a multi-file study.
- **Randomized:** Variables generated by CAI surveys produced by one or more random number variables together with a response variable, e.g, random variable X which could equal 1 or 2 (at random) which in turn would control whether Q.23 is worded "men" or "women", e.g., would you favor helping [men/women] laid off from a factory obtain training for a new job?
- **Other:** Variables which do not fit easily into any of the categories listed above, e.g., a group of variables whose documentation is in another language.

  ❖ The "*varGrp*" attribute is used to reference all the subsidiary variable groups which nest underneath the current "*varGrp*". This allows for encoding of a hierarchical structure of variable groups.
  ❖ The attribute "*name*" provides a name, or short label, for the group.
  ❖ The "*sdatrefs*" are summary data description references that record the ID values of all elements within the summary data description section of the Study Description that might apply to the group. These elements include: time period covered, date of collection, nation or country, geographic coverage, geographic unit, unit of analysis, universe, and kind of data.
  ❖ The "methrefs" are methodology and processing references which record the ID values of all elements within the study methodology and processing section of the Study Description which might apply to the group. These elements include information on data collection and data appraisal (e.g., sampling, sources, weighting, data cleaning, response rates, and sampling error estimates).
  ❖ The "*pubrefs*" attribute provides a link to publication/citation references and records the ID values of all citations elements within Section 2.5 or Section 5.0 that pertain to this variable group.
  ❖ The "*access*" attribute records the ID values of all elements in Section 2.4 of the document that describe access conditions for this variable group.

➢ <u>Var</u>
Listing of variables that make up the Variable Group

4.1.1   <labl >   Label

- Optional
- Repeatable
- Attributes: <u>ID, xml:lang, source</u>, level, vendor, country, sdatrefs

Description: A short description of the parent element, that is, the group. In the variable label, the length of this phrase may depend on the statistical analysis system used (e.g., some versions of SAS permit 40-character labels, while some versions of SPSS permit 120 characters), although the DDI itself imposes no restrictions on the number of characters allowed. A "level" attribute is included to permit coding of the level to which the label applies, i.e. record group, variable group, variable, category group, category, nCube group, nCube, or other study-related materials. The "vendor" attribute was provided to allow for specification of different labels for use with different vendors' software. The attribute "country" allows for the denotation of country-specific labels. The "sdatrefs" attribute records the ID values of all elements within the Summary Data Description section of the Study Description that might apply to the label. These elements include: time period covered, date of collection, nation or country, geographic coverage, geographic unit, unit of analysis, universe, and kind of data.

<odesi>
  format

Formatting Note:
   Only the first word in a phrase should be capitalized.

Example 1:
   <varGrp ID="**VG1F1**" type="**subject**" var="**V1 V2 V3**">
   <labl>**Identification variables**</labl>
    </varGrp>

Example 2:
   <varGrp ID="**VG2F1**" type="**subject**" var ="**V4 V5**">
   <labl>**Demographic variables - household**</labl>
   </varGrp>

Example 3:
   <varGrp ID="**VG3F1**" type="**subject**" var="**V6 V7 V8 V9 V10 V11 V12 V13 V14**">
   <labl>**Household smoking**</ labl>
   </varGrp>

4.1.6   <notes>   Notes and Comments

- Optional
- Repeatable
- Attributes: <u>ID, xml:lang, source</u>, type, subject, level, resp, sdatrefs

Description: For clarifying information/annotation regarding the parent element. The attributes for notes permit a controlled vocabulary to be developed ("type" and "subject"), indicate the "level" of the DDI to which the note applies (study, file, variable, etc.), and identify the author of the note ("resp").

Example:
   <dataDscr><notes>**The variables in this study are identical to earlier waves.**
   </notes>
   </dataDscr>

4.3   <var>   Variable

- Optional
- Repeatable
- Attributes: <u>ID, xml:lang, source</u>, name, wgt, wgt-var, weight, qstn, files, vendor, dcml, intrvl, rectype, sdatrefs, methrefs, pubrefs, access, aggrMeth, measUnit, scale, origin, nature, additivity, temporal, geog, geoVocab, catQnty

Description: This element describes all of the features of a single variable in a social science data file. There are a number of attributes within this tag that allows us to specify items such as the name of the variable, what the weight variable is, etc….

Attributes within the <var>
- <u>ID</u>
  To uniquely identify the Variable

- <u>Name</u>
  Name of the variable

- <u>Files</u>
  Indicates the file where the variable can be found

- <u>Wgt</u>
  Indicates whether the variable is the weight variable or not

- <u>Dcml</u>
  Indicates the number of decimal places that are held in the variable

- <u>Intrvl</u>
  Indicates whether the variable is continuous or discrete, ordinal or nominal.

- <u>Wgt-var</u>
  References the weight variable(s) for this variable.

Example 1:
   <var ID="**V1**" name="**CASEID**" files="**F1**" dcml="**0**" intrvl="**contin**">
Explanation of Example 1:
Variable V1 has a name of Caseid, has 0 decimals and is a continuous variable.

Example 2:
   <var ID="**V2**" name="**WEIGHT**" wgt="**wgt**" files="**F1**" dcml="**0**" intrvl="**discrete**">
 Explanation of Example 2:
Variable V2 has the name WEIGHT and has been designated as the Weight variable.

Example 3:
   <var ID="**V8**" name="**AGE1524**" wgt-var="**V154**" files="**F1**" dcml="**0**" intrvl="**contin**">
Explanation of Example 3:
Variable V8 has the name AGE1524 and the appropriate weight to use is V154.

4.3.1   <location>   Location

- Optional
- Repeatable
- Attributes: <u>ID, xml:lang, source</u>, StartPos, EndPos, width, RecSegNo, fileid, locMap

Description: This is an empty element containing only the attributes listed below. Attributes include "StartPos" (starting position of variable), "EndPos" (ending position of variable), "width" (number of columns the variable occupies), "RecSegNo" (the record segment number, deck or card number the variable is located on), and "fileid" (an IDREF link to the fileDscr element for the file that this location is within). The fileid is necessary in cases where the same variable may be coded in two different files, e.g., a logical record length type file and a card image type file. Note that if there is no width or ending position, then the starting position should be the ordinal position in the file, and the file would be described as free-format.

Attributes within the <location>

- <u>StartPos</u>
  Indicates the starting position for the variable in the Datafile

- <u>EndPos</u>
  Indicates the ending position for the variables in the Datafile

- <u>Width</u>
  Indicates the width of the variable in the Datafile.

Note:
  See Appendix C for more information on this tag.

Example 1:
  <location StartPos="**9**" EndPos="**9**" width="**1**" />
  Explanation of Example 1:
  The variable starts in column 9 and ends in column 9 with a width of 1.

Example 2:
  <location StartPos="**10**" EndPos="**17**" width="**8**" />
  Explanation of Example 2:
  The variable starts in column 10 and ends in column 17 with a width of 8.

4.3.2   <labl>   Label

- Optional
- Repeatable
- Attributes: <u>ID, xml:lang, source</u>, level, vendor, country, sdatrefs

Description: A short description of the parent element. In the variable label, the length of this phrase may depend on the statistical analysis system used (e.g., some versions of SAS permit 40-character labels, while some versions of SPSS permit 120 characters), although the DDI itself imposes no restrictions on the number of characters allowed. A "level" attribute is included to permit coding of the level to which the label applies, i.e. record group, variable group, variable, category group, category, nCube group, nCube, or other study-related materials. The "vendor" attribute was provided to allow for specification of different labels for use with different vendors' software. The attribute "country" allows for the denotation of country-specific labels. The "sdatrefs" attribute records the ID values of all elements within the Summary Data Description section of the Study Description that might apply to the label. These elements include: time period covered, date of collection, nation or country, geographic coverage, geographic unit, unit of analysis, universe, and kind of data.

Example 1:
   <labl>**Household income**</labl>

Example 2:
   <labl>**Gender**</labl>

Example 3:
   <labl>**Province of residence** </labl>

Example 4:
   <labl>**Internet use – content concern** </labl>

4.3.8   <qstn>   Question

- Optional
- Repeatable
- Attributes: <u>ID, xml:lang, source</u>, qstn, var, seqNo, sdatrefs

Description: The question element may have mixed content. The element itself may contain text for the question, with the sub-elements being used to provide further information about the question. Alternatively, the question element may be empty and only the sub-elements used. The element has a unique question ID attribute which can be used to link a variable with other variables where the same question has been asked. This would allow searching for all variables that share the same question ID perhaps because the questions was asked several times in a panel design.

Note:
 This information can be obtained from the questionnaire.

4.3.8.1   <preQTxt>   PreQuestion Text

- Mandatory
- Not Repeatable
- Attributes: <u>ID, xml:lang, source</u>

Description: Text describing a set of conditions under which a question might be asked.

Note 1:
    See Appendix C for more information on this tag.

Note 2:
    This information can be obtained from the questionnaire.

Example:
    <preQTxt> **This survey deals with various aspects of your health. I'll be asking about such things as physical activity, social relationships and health status. By health, we mean not only the absence of disease or injury but also physical, mental and social well-being**. </preQTxt>

4.3.8.2   <qstnLit>   Literal Question

- Mandatory
- Not Repeatable
- Attributes: <u>ID, xml:lang, source</u>, sdatrefs

Description: Text of the actual, literal question asked.

Note 1:
    See Appendix C for more information on this tag.

Note 2:
  This information can be obtained from the questionnaire.

Example:
    <qstnLit> **During June to August, when you were at work, how much time each day (on average) were you in the sun?** </qstnLit>

4.3.8.3   <postQTxt>   PostQuestion Text

- Mandatory
- Not Repeatable
- Attributes: <u>ID, xml:lang, source</u>

Description: Text describing what occurs after the literal question has been asked.

Note 1:
    See Appendix C for more information on this tag.

Note 2:
  This information can be obtained from the questionnaire.

Example:
    <postQTxt>**Go to next module**. </postQTxt>

4.3.8.4   <forward>   Forward Progression

- Optional
- Repeatable
- Attributes: <u>ID, xml:lang, source</u>, qstn

Description: Contains a reference to IDs of possible following questions. The "qstn" IDREFS may be used to specify the question IDs.

  Note 1:
     See Appendix C for more information on this tag.

  Note 2:
     This information can be obtained from the questionnaire.

  Example 1:
     <var><qstn><forward qstn="**Q120 Q121 Q122 Q123 Q124**"> **If yes, please ask questions 120-124**.</forward></qstn></var>

4.3.8.5   <backward>   Backflow

- Optional
- Repeatable
- Attributes: <u>ID, xml:lang, source</u>, qstn

Description: Contains a reference to IDs of possible preceding questions. The "qstn" IDREFS may be used to specify the question IDs.

  Note 1:
     See Appendix C for more information on this tag.

  Note 2:
     This information can be obtained from the questionnaire.

  Example:
     <var><qstn><backward qstn="**Q12 Q13 Q14 Q15**">**For responses on a similar topic, see questions 12-15.**</backward></qstn> </var>

4.3.8.6   <ivuInstr>   Interviewer Instructions

- Mandatory
- Not Repeatable
- Attributes: <u>ID, xml:lang, source</u>

Description: Specific instructions to the individual conducting an interview.

  Note 1:
     See Appendix C for more information on this tag.

  Note 2:
      This information can be obtained from the questionnaire.

Example:
    <ivuInstr>**Read list. Mark all that apply.** </ivuInstr>

4.3.9   <valrng>   Range of Valid Data Values

- Optional
- Repeatable
- Attributes: <u>ID, xml:lang, source</u>

Description: Values for a particular variable that represent legitimate responses.

Note:
    See Appendix C for more information on this tag.

Example 1:
    <valrng><range min="**1**" max="**3**" /></valrng>

Example 2:
    <valrng><item VALUE="**1**" /><item VALUE="**2**" /><item VALUE="**3**" /></valrng>

4.3.9.2   <notes>   Notes and Comments

- Optional
- Repeatable
- Attributes: <u>ID, xml:lang, source</u>, type, subject, level, resp, sdatrefs

Description: For clarifying information/annotation regarding the parent element. The attributes for notes permit a controlled vocabulary to be developed ("type" and "subject"), indicate the "level" of the DDI to which the note applies (study, file, variable, etc.), and identify the author of the note ("resp").

4.3.10   <invalrng>   Range of Invalid Data Values

- Optional
- Repeatable
- Attributes: <u>ID, xml:lang, source</u>

Description: Values for a particular variable that represent missing data, not applicable responses, etc.

Note:
    See Appendix C for more information on this tag.

Example:
    <invalrng> <range UNITS="INT" min="**98**" max="**99**"> </range>
    <key> **98 DK**
    99 Inappropriate
    </key> </invalrng>

4.3.10.2   \<notes\>   Notes and Comments

- Optional
- Repeatable
- Attributes: <u>ID, xml:lang, source</u>, type, subject, level, resp, sdatrefs

Description: For clarifying information/annotation regarding the parent element. The attributes for notes permit a controlled vocabulary to be developed ("type" and "subject"), indicate the "level" of the DDI to which the note applies (study, file, variable, etc.), and identify the author of the note ("resp").

4.3.11   \<undocCod\>   List of Undocumented Codes

- Optional
- Repeatable
- Attributes: <u>ID, xml:lang, source</u>

Description: Values whose meaning is unknown.

Note:
    See Appendix C for more information on this tag.

Example:
    \<var\>\<undocCod\>**Responses for categories 9 and 10 are unavailable**.\</undocCod\>\</var\>

4.3.12   \<universe\>   Universe

- Optional
- Repeatable
- Attributes: <u>ID, xml:lang, source</u>, level, clusion

Description: The group of persons or other elements that are the object of research and to which any analytic results refer. Age, nationality, and residence commonly help to delineate a given universe, but any of a number of factors may be involved, such as sex, race, income, veteran status, criminal convictions, etc. The universe may consist of elements other than persons, such as housing units, court cases, deaths, countries, etc. In general, it should be possible to tell from the description of the universe whether a given individual or element (hypothetical or real) is a member of the population under study. A "level" attribute is included to permit coding of the level to which universe applies, i.e., the study level, the file level (if different from study), the record group, the variable group, the nCube group, the variable, or the nCube level. The "clusion" attribute provides for specification of groups included (I) in or excluded (E) from the universe. If all the variables nCubes described in the data documentation relate to the same population, e.g., the same set of survey respondents, this element would be unnecessary at data description level. In this case, universe can be fully described at the study level.

Note 1:
    See Appendix C for more information on this tag.

Note 2:
    This is the universe for each individual question.

Example:
    <universe clusion="I">**Included: Individuals 15-19 years of age.** </universe>
    <universe clusion="E">**Excluded: Individuals younger than 15 and older than 19**
    years of age.</universe>

4.3.13   <TotlResp>   Total Responses

- Optional
- Not Repeatable
- Attributes: <u>ID, xml:lang, source</u>

Description: The number of responses to this variable. This element might be used if the number of responses does not match added case counts. It may also be used to sum the frequencies for variable categories.

Example 1:
    <var><TotlResp>**1,056**</TotlResp></var>

Example 2:
    <var><TotlResp>**There are only 725 responses to this question since it was not asked in Tanzania.**</TotlResp></var>

4.3.14   <sumStat>   Summary Statistics

- Optional
- Repeatable
- Attributes: <u>ID, xml:lang, source</u>, wgtd, wgt-var, weight, type

Description: One or more statistical measures that describe the responses to a particular variable and may include one or more standard summaries, e.g., minimum and maximum values, median, mode, etc. The attribute "wgtd" indicates whether the statistics are weighted or not. The "weight" attribute is an IDREF(S) to the weight element(s) in the study description. The attribute "type" denotes the type of statistics being shown: mean, median, mode, valid cases, invalid cases, minimum, maximum, or standard deviation.

Example 1:
    <var><sumStat type="min">**0**</sumStat></var>

Example 2:
    <var><sumStat type="max">**9**</sumStat></var>

Example 3:
    <var><sumStat type="**median**">4</sumStat></var>

4.3.16   <stdCatgry>   Standard Categories

- Optional
- Repeatable
- Attributes: <u>ID, xml:lang, source</u>, date, URI

Description: Standard category codes used in the variable, like industry codes, employment codes, or social class codes. The attribute "date" is provided to indicate the version of the code in place at the time of the study. The attribute "URI" is provided to indicate a URN or URL that can be used to obtain an electronic list of the category codes.

Example:
    <var><stdCatgry date="**1981**" source="producer">**U. S. Census of Population and Housing, Classified Index of Industries and Occupations** </stdCatgry></var>

4.3.18   <catgry>   Category

- Optional
- Repeatable
- Attributes: <u>ID, xml:lang, source</u>, missing, mistype, country, sdatrefs,excls

Description: A description of a particular response. The attribute "missing" indicates whether this category group contains missing data or not. The attribute "missType" is used to specify the type of missing data, e.g. inap, don't know, no answer, etc. The attribute "country" allows for the denotation of country-specific category values. The "sdatrefs" attribute records the ID values of all elements within the summary data description that apply to this category. The exclusiveness attribute ("excls") should be set to "false" if the category can appear in more than one place in the classification hierarchy.

4.3.18.1   <catValu>   Category Value

- Optional
- Not Repeatable
- Attributes: <u>ID, xml:lang, source</u>

Description: The explicit response.

Note:
    See Appendix C for more information on this tag.

Example 1:
    <var><catgry missing="**Y**" missType="**inap**"><catValu>**9**</catValu></catgry></var>

Example 2:
    <labl>**Always**</labl>
    <catValu>**24**</catValu>
    Explanation of Example 2:
    The value for the given category of 'Always' in Var X is 24.

4.3.18.2   <labl>   Label

- Optional
- Repeatable
- Attributes: <u>ID, xml:lang, source</u>, level, vendor, country, sdatrefs

Description: A short description of the response. In the variable label, the length of this phrase may depend on the statistical analysis system used (e.g., some versions of SAS permit 40-character labels, while some versions of SPSS permit 120 characters). Although the DDI itself imposes no restrictions on the number of characters allowed. A "level" attribute is included to permit coding of the level to which the label applies, i.e. record group, variable group, variable, category group, category, nCube group, nCube, or other study-related materials. The "vendor" attribute was provided to allow for specification of different labels for use with different vendors' software. The attribute "country" allows for the denotation of country-specific labels. The "sdatrefs" attribute records the ID values of all elements within the Summary Data Description section of the Study Description that might apply to the label. These elements include: time period covered, date of collection, nation or country, geographic coverage, geographic unit, unit of analysis, universe, and kind of data.

Note:
  See Appendix C for more information on this tag.

Example 1:
  <labl>**Always**</labl>
  Explanation of Example 1:
  The value label for the category in Var X is "Always"

Example 2:
  <labl>**Not in the labour force**</labl>
  Explanation of Example 2:
  The value label for the category in Var X is "Not in the labour force".

4.3.18.4   <catStat>   Category Group Statistics

- Optional
- Repeatable
- Attributes: <u>ID, xml:lang, source</u>, type, URI, methrefs, wgtd, wgt-var, weight, sdatrefs

Description: May include frequencies, percentages, or cross-tabulation results which define the category; often appears in a table. This field can contain one of the following: 1) textual information (e.g. PCDATA); or 2) non-parseable character data (e.g. the statistics); or 3) some other form of external information (table, image, etc.). In case 1, the tag can be used to mark up character data; tables can also be included in the actual markup. In cases 2 or 3, the element can be left empty and the "URI" attribute used to refer to the external object containing the information. The attribute "type" indicates the type of statistics presented – frequency, percent, or crosstabulation.

Attributes within the <catStat>

  <u>Type</u>
  The attribute "type" refers to "frequency", "percent", or "crosstab".

Note:
  See Appendix C for more information on this tag.

Example:
  <catStat type="**freq**">**16385**</catStat>
  Explanation of Example:
  A frequency was calculated with a value of 16385.

4.3.19   <codInstr>   Coder Instructions

- Optional
- Repeatable
- Attributes: <u>ID, xml:lang, source</u>

Description: Any special instructions to those who converted information from one form to another for a particular variable. This might include the reordering of numeric information into another form or the conversion of textual information into numeric information.

Note:
     See Appendix C for more information on this tag.

Example:
     <var><codInstr>**Use the standard classification tables to present responses to the question: "What is your occupation?" into numeric codes.**</codInstr></var>

4.3.22   <derivation>   Derivation

- Optional
- Not Repeatable
- Attributes: <u>ID, xml:lang, source</u>, var

Description: Used only in the case of a derived variable, this element provides both a description of how the derivation was performed and the command used to generate the derived variable, as well as a specification of the other variables in the study used to generate the derivation. The "var" attribute provides the ID values of the other variables in the study used to generate this derived variable.

4.3.22.1   <drvdesc>   Derivation Description

- Optional
- Not Repeatable
- Attributes: <u>ID, xml:lang, source</u>

Description: A textual description of the way in which this variable was derived.

Note:
     See Appendix C for more information on this tag

Example 1:
     <title>Canadian Community Health Survey, 2005: Cycle 3.1, Main File </title>
     <var><derivation><drvdesc> **DHHnDHSZ "Household Size" This variable indicates the number of people living within a household.   This variable is derived by sorting the household roster dataset by SAMPLEID and PERSONID and by counting the number of PERSONID's within each SAMPLEID.**
     </drvdesc></derivation></var>

Example 2:
     <var><derivation><drvdesc> **VAR215.01 "Outcome of first pregnancy" (1988 NSFG=VAR611 PREGOUT1) If R has never been pregnant (VAR203 PREGNUM EQ 0) then OUTCOM01 is blank/inapplicable. Else, OUTCOM01 is transferred from VAR225 OUTCOME for R's 1st pregnancy.** </drvdesc></derivation></var>

4.3.22.2   \<drvcmd\>   Derivation Command

- Optional
- Not Repeatable
- Attributes: <u>ID, xml:lang, source</u>, syntax

Description: The actual command used to generate the derived variable. The "syntax" attribute is used to indicate the command language employed (e.g., SPSS, SAS, Fortran, etc.)

Note:
See Appendix C for more information on this tag

Example:
\<var\>\<derivation\>\<drvcmd syntax="SPSS" \>**RECODE V1 TO V3 (0=1) (1=0) (2=-1) INTO DEFENSE WELFARE HEALTH.** \</drvcmd\>\</derivation\>\</var\>

4.2.23   \<varFormat\>   Variable Format

- Optional
- Not Repeatable
- Attributes: <u>ID, xml:lang, source</u>, type, formatname, schema, category, URI

Description: The technical format of the variable in question. Attributes for this element include: "type," which indicates if the variable is character or numeric; "formatname," which in some cases may provide the name of the particular, proprietary format actually used; "schema," which identifies the vendor or standards body that defined the format (acceptable choices are SAS, SPSS, IBM, ANSI, ISO, XML-data or other); "category," which describes what kind of data the format represents, and includes date, time, currency, or "other" conceptual possibilities; and "URI," which supplies a network identifier for the format definition.

Note:
See Appendix C for more information on this tag

Example 1:
   \<var\>\<varFormat type="numeric" schema="SAS" formatname="DATE" category="date" \>**The number in this variable is stored in the form 'ddmmmyy' in SAS format.** \</varFormat\>\</var\>

Example 2:
   \<var\>\<varFormat type="numeric" formatname="date.iso8601" schema="XML-Data" category="date" URI="**http://www.w3.org/TR/1998/NOTE-XML-data/**">19541022 \</varFormat\>\</var\>

4.3.26   \<notes\>   Notes and Comments

- Optional
- Repeatable
- Attributes: <u>ID, xml:lang, source</u>, type, subject, level, resp, sdatrefs

Description: For clarifying information/annotation regarding the parent element. The attributes for notes permit a controlled vocabulary to be developed ("type" and "subject"), indicate the "level" of the DDI to which the note applies (study, file, variable, etc.), and identify the author of the note ("resp").

Example:
    <dataDscr><notes>**The variables in this study are identical to earlier waves.**
    </notes>
    </dataDscr>

4.4   <notes>   Notes and Comments

- Optional
- Repeatable
- Attributes: <u>ID, xml:lang, source</u>, type, subject, level, resp, sdatrefs

Description: For clarifying information/annotation regarding the parent element. The attributes for notes permit a controlled vocabulary to be developed ("type" and "subject"), indicate the "level" of the DDI to which the note applies (study, file, variable, etc.), and identify the author of the note ("resp").

SECTION 5.0   OTHER DOCUMENTATION   <otherMat>

This section allows for the inclusion of other materials that are related to the study as identified and labeled by the DTD users (encoders). The materials may be entered as PCDATA (ASCII text) directly into the document (through use of the "txt" element). This section may also serve as a "container" for other machine-readable materials such as data definition statements by providing a brief description of the study-related materials accompanied by the attributes "type" and "level" defining the material further. The "URI" attribute may be used to indicate the location of the other study-related materials.

Other Study-Related Materials may include: questionnaires, coding notes, SPSS/SAS/STATA setups (and others), user manuals, continuity guides, sample computer software programs, glossaries of terms, interviewer/project instructions, maps, database schema, data dictionaries, show cards, coding information, interview schedules, missing values information, frequency files, variable maps, etc.

Note: In this section, links are added to other materials. These links go to the location of the material on the website for **your** workplace.

Tags and corresponding DTD numbers are from DDI Version 2.0 codebook.

Source: DDI Codebook
http://www.icpsr.umich.edu/DDI/dtd/version2-1-all.html
http://www.icpsr.umich.edu/DDI/dtd/version2-1-tree.html

DTD Numbers	Tags
5.0	<otherMat>
5.1	<labl>
5.2	<txt>
5.3	<notes>

Description of Tags and Working Examples

5.0   <otherMat>   Other Study-Related Materials

- Optional
- Repeatable
- Attributes: <u>ID, xml:lang, source</u>, type, level, URI

Description: Other materials related to the study. This section allows for the inclusion of other materials that are related to the study as identified and labeled by the DTD/Schema users (encoders). The' materials may be entered as PCDATA (ASCII text) directly into the document (through use of the "txt" element). This section may also serve as a "container" for other electronic materials such as setup files by providing a brief description of the study-related materials accompanied by the attributes "type" and "level" defining the material further. The "URI" attribute may be used to indicate the location of the other study-related materials.

Other Study-Related Materials may include: questionnaires, coding notes, SPSS/SAS/Stata setup files (and others), user manuals, continuity guides, sample computer software programs, glossaries of terms, interviewer/project instructions, maps, database schema, data dictionaries, show cards, coding information, interview schedules, missing values information, frequency files, variable maps, etc.

Note that Section 2.5, Other Study Description Materials, should be used for materials that are primarily descriptions of the content and use of the study, such as appendices, sampling information, weighting details, methodological and technical details, publications based upon the study content, related studies or collection of studies, etc. This section, 5.0 Other Study-Related Materials, is intended to include or to link to materials used in the production of the study or useful in the analysis of the study.

The "level" attribute is used to clarify the relationship of the other materials to components of the study. Suggested values for level include specifications of the item level to which the element applies: e.g., level=datafile; level=studydsc; level=study. The URI attribute need not be used in every case; it is intended for capturing references to other materials separate from the codebook itself. In Section 5, Other Material is recursively defined.

Example 1:
   <otherMat
   URI="http://tdr.uoguelph.ca/DATA/CONSUM/SHS/2001/DOCS/shs2001que.pdf">
   </otherMat>

Example 2:
   <otherMat
   URI="http://www.statcan.ca/english/Dli/Metadata/shs/2005/shs2005que.pdf">
   </otherMat>

5.1   <labl>   Label

- Optional
- Repeatable
- Attributes: <u>ID, xml:lang, source</u>, level, vendor, country, sdatrefs

Description: Short description of the other material. A "level" attribute is included to permit coding of the level to which the label applies, i.e., the study level, the file level (if different from study), the record group, the variable group, or the variable level. Vendor attribute provided to allow for specification of different labels for use with different vendors' software.

Example 1:
   <labl>**Questionnaire - PDF**</labl>

Example 2:
   **<otherMat level="study"
   URI="http://nesstar.tdr.uoguelph.ca/HEALTH/CCHS/2005/DOCS/CVTAB_E.pdf">
   <labl> Approximate Sampling Variability Tables - Main File – PDF </labl>
   </otherMat>**

5.2   <txt>   Text

- Optional
- Not Repeatable
- Attributes: <u>ID, xml:lang, source</u>, level, sdatrefs

Description: Lengthier description of other material. A "level" attribute is included to permit coding of the level to which the text applies, i.e., the study level, the file level (if different from study), the record group, the variable group, or the variable level.

Example 1:
   < otherMat level="**study**"
   URI="**http://www.adobe.com/prodindex/acrobat/readstep.html**"><labl>**CLICK HERE
   to download Adobe Acrobat**</labl>
    <txt>**The documentation for some surveys is produced in a variety of formats. In
    the case of PDF files, you will require Adobe Acrobat to access the files**</txt>
   </otherMat>

Example 2:
   <otherMat type="**Users Guide:**" level="**study**" URI="**http://tdr.uoguelph.ca/DATA/
   CONSUM/SHS/2001/DOCS/shs2001gid.pdf**"> <labl>**Users Guide - PDF**</labl>
   </otherMat>

5.3   <notes>   Notes and Comments

- Optional
- Repeatable
- Attributes: <u>ID, xml:lang, source</u>, type, subject, level, resp, sdatrefs

Description: Used to indicate additional information about the other material. "Notes" sections appear in several places in the DTD. The attributes for notes permit a controlled vocabulary to be developed (type and subject), the level of the DTD to which the note refers to be identified (study, file, variable, etc.), and the author of the note to be indicated (resp).

Appendix A: Glossary

Bootstrap File
- A type of data file which contains Bootstrap weights

Bootstrap Weights
- Mimics taking several new samples from the population.
- Used to compute the precision of an estimation – for example the CV of the number of smokers estimation.

CDATA
- CDATA means character data.
- CDATA is text that will NOT be parsed by a parser. Tags inside the text will NOT be treated as markup and entities will not be expanded.

DDI
- "The Data Documentation Initiative (DDI) is an effort to establish an international XML-based standard for the content, presentation, transport, and preservation of documentation for datasets in the social and behavioural sciences. Documentation, sometimes called metadata (data about data), constitutes the information that enables the effective, efficient, and accurate use of those datasets." (http://www.ddialliance.org/codebook/index.html)

DDI Mark-up Tools
- There are many tools available to mark-up DDI. NESSTAR is merely one of them. See this site for more tools. http://www.ddialliance.org/DDI/related/tools.html

DTD
- "The DDI is currently expressed as an XML Document Type Definition (DTD). The DTD defines all of the elements and attributes of social science technical documentation and the relationships among the elements and attributes." See Appendix C for more information. (http://www.ddialliance.org/codebook/index.html)
- XML uses a DTD or an XML schema to describe the data
- The DDI uses both a DTD and follows an XML schema

Infrastructure
- This field is filled in if an agency/department has given money or any other type of support to help with any aspect of the survey.

Metadata
- The documentation that accompanies and assists users in the interpretation of different types of data. The information usually includes: the description of the methodology; the definition of the variables; and any other information related to the data.
- The data describing context, content and structure of records and their management through time.

Nesstar
- A tool used to mark-up DDI.

ODESI
- Ontario Data Documentation, Extraction Service and Infrastructure Initiative (ODESI)
- Provides Ontario university researchers with access to datasets in a web-based data extraction system delivered through Scholars Portal.
- ODESI is jointly funded by OntarioBuys and the OCUL community.

OntarioBuys
- http://www.fin.gov.on.ca/ontariobuys/
- "OntarioBuys is a program of the Ontario Ministry of Finance that provides funding and advice to the province's broader public sector (BPS) partners – principally, hospitals, school boards, colleges and universities – to help them modernize their supply chains and other back office processes." (taken from the OntarioBuys website)

PCDATA
- PCDATA means parsed character data.
- Think of character data as the text found between the start tag and the end tag of an XML element.
- PCDATA is text that WILL be parsed by a parser. The text will be examined by the parser for entities and markup.
- Tags inside the text will be treated as markup and entities will be expanded.
- However, parsed character data should not contain any &, <, or > characters; these need to be represented by the & < and > entities, respectively.

PUMF File
- Acronym for Public Use Microdata file
- Contains data that has been vetted by Data Producers to ensure that tabulations will not allow identification of survey participants
- Researchers can analyse any variable in the PUMF and can choose which crosstabulations they will use rather than having to use the tables given in aggregate format.

Scholars Portal
- http://www.scholarsportal.info/index.html

Synthetic File
- A type of data file
- Also known as a dummy file
- Provides the full variable structure of the master file but does not contain any real cases, therefore it can never be used to compile actual statistics.
- Used to assist researchers to create program files at local Data Centres that can then be used to access Master Files in a Research Data Centre or through Remote Job Submission.

XML Briefer
- XML stands for e**X**tensible **M**arkup **L**anguage.
- XML has been designed to describe data.
- XML is similar in format to HTML
- XML describes data and focuses on what data is
- HMTL displays data and focuses on how data looks

XML Schema
- XML uses a DTD (Document Type Definition) or an XML schema to describe the data

XML Tag
- <dataDscr> - is an example of an XML tag
- XML tags are case sensitive
- Need an opening and closing tag
    - < dataDscr> is an opening tag
    - </dataDscr> is a closing tag

Appendix B: Nesstar Related Information

The following section lists information specific to using DDI with Nesstar.

A. If using Nesstar Publisher, the following fields are filled in automatically.

1.1.3.5	<software>	Software used in Production
2.1.3.5	<software>	Software used in Production
Section 3.0	<fileDscr>	Data Files Description
Except for **3.1.2, **3.1.8**, **3.1.12**		
4.3.1	<location>	Location
4.3.9	<valrng>	Range of Valid Data Values
4.3.18.1	<catValu>	Category Value
4.3.18.2	<labl>	Label
4.3.18.4	<catStat>	Category Group Statistics

B. Note re: the data file:
    When a .sav file is pulled into NESSTAR, it becomes an NSDstat file. Later on, if this file is exported into SPSS format, the new .sav file will be different from the original .sav file.

C. There are a few tags that are labelled differently in Nesstar than in DDI 2.x.
    This document is based on Nesstar using DDI 2.x
    - Section 2:
        Unit of Analysis:
            DDI - **2.2.3.8**; Nesstar - **2.2.3.6**
        Universe:
            DDI - **2.2.3.9**; Nesstar - **2.2.3.7**
         Kind of Data:
            DDI - **2.2.3.10**; Nesstar - **2.2.3.8**
    -
        Section 4:
         Question:
            DDI - **4.3.8**; Nesstar - **4.2.8**
         Pre-question text:
            DDI - **4.3.8.1**; Nesstar - **4.2.8.1**
         Literal question:
            DDI - **4.3.8.2**; Nesstar - **4.2.8.2**
         Post-question Text:
            DDI - **4.3.8.3**; Nesstar - **4.2.8.3**
         Interviewer Instructions:
             DDI - **4.3.8.6**; Nesstar - **4.2.8.6**
        Universe
            DDI – **4.3.12**; Nesstar – **4.2.12**
        Notes and Comments
            DDI – **4.3.26**; Nesstar – **4.2.24**

D. Many of the attributes listed in the DDI document are not available in the Nesstar Publisher.

E. Tag 2.1.5.1   <serName>   Name of Series

- Nesstar does not allow you to input the abbreviation of the survey series name for tag 2.1.5.1.

F. Using a template is a real time-saver. For more information on templates, please contact the <odesi> developers at odesi@uoguelph.ca

G. There are some tags which cannot be added if you are using Nesstar. The following list is a partial one.

4.3.22.1   <drvdesc>   Derivation Description
4.3.22.2   <drvcmd>   Derivation Command

Appendix C: Statistics Canada Related Information

1. Statistics Canada (STC) Sources to Obtain Tag Information

- There are many different STC resources that can be used to obtain information to fill in tags listed in the BPD. Some of the resources include:
    - survey documention (user guide, questionnaire, codebook);
    - STC Online Catalogue;
    - BiblioCat (STC's library catalogue); and
    - IMDB.

2. The following tags have information related to surveys from Statistics Canada.

1.1.5.2   \<serInfo\>   Series Information
Note:
If it is a Statistics Canada survey that is being marked up, this information is usually available from the Product Page at Statistics Canada. It may also be available in the abstract or the User's Guide.

2.1.5.2   \<serInfo\>   Series Information
Note:
If it is a Statistics Canada survey that is being marked up, this information is usually available from the Product Page at Statistics Canada. It may also be available in the abstract or the User's Guide.

2.2.1.1   \<keyword\>   Keywords
Note:
If there are keywords listed on the survey description page on the Statistics Canada website, use these keywords for this tag.

2.2.1.2   \<topcClas\>   Topic Classification
Note:
If there are topic classifications listed on the survey description page on the Statistics Canada website, use them for this tag.

3. Check the Microdata User Guide or the Codebook or the Questionnaire or other Statistics Canada Sources for information related to the following tags.

Please note: this is a partial list.

2.1.2.2	Other ID/Acknowledgements
2.1.5.1	Name of Series
2.2.3.1	Time Period Covered
2.2.3.2	Date of Collection
2.2.3.3	Country
2.2.3.4	Geographic Coverage
2.2.3.5	Geographic Unit
2.2.3.9	Universe
2.2.3.10	Kind of Data
2.2.4	Notes
2.3.1.1	Time Method
2.3.1.2	Data Collector
2.3.1.3	Frequency of Data Collection
2.3.1.4	Sampling Procedure
2.3.1.6	Mode of Collection
2.3.1.9	Characteristics of the Data Collection Situation
2.3.1.12	Weighting

4. Tag 2.1.3.2   Copyright

This tag should not contain information about the division that conducted the survey.

> *Example:*
> <titl>**Canadian Tobacco Use Monitoring Survey, 2004: Cycle 1, Household File**</titl>
> <copyright>**Copyright © Statistics Canada, 2005**</copyright>

Appendix D: Data Documentation Initiative (DDI)

This appendix summarizes the key objectives of the DDI. Version 1.0 of the DTD was published March 24, 2000. Since that time, several enhancements have been made; the most recent stable version of the specification is Version 2.1.

All the information for this appendix was taken from the DDI site. For more detailed information about the DDI, please go to: http://www.ddialliance.org/codebook/index.html

The DDI facilitates:

- Interoperability. Codebooks marked up using the DDI specification can be exchanged and transported seamlessly, and applications can be written to work with these homogeneous documents.

- Richer content. The DDI was designed to encourage the use of a comprehensive set of elements to describe social science datasets as completely and as thoroughly as possible, thereby providing the potential data analyst with broader knowledge about a given collection.

- Single document - multiple purposes. A DDI codebook contains all of the information necessary to produce several different types of output, including, for example, a traditional social science codebook, a bibliographic record, or SAS/SPSS/Stata data definition statements. Thus, the document may be repurposed for different needs and applications. Changes made to the core document will be passed along to any output generated.

- On-line subsetting and analysis. Because the DDI markup extends down to the variable level and provides a standard uniform structure and content for variables, DDI documents are easily imported into on-line analysis systems, rendering datasets more readily usable for a wider audience.

- Precision in searching. Since each of the elements in a DDI-compliant codebook is tagged in a specific way, field-specific searches across documents and studies are enabled. For example, a library of DDI codebooks could be searched to identify datasets covering protest demonstrations during the 1960s in specific states or countries.

Appendix E: <odesi> Related Information

Listserv
- To keep in touch with new developments, post questions, …
- http://odesi.uoguelph.ca/wiki/index.php/ODESI_listserve

Website
- http://odesi.scholarsportal.info.proxy.library.carleton.ca/webview/
- Please note: Only Ontario Universities are allowed to access this site.

Wiki
- This wiki is open to all.
- http://odesi.uoguelph.ca/wiki/index.php/Main_Page

Tags specific to <odesi>

    1.1.3.6   <fundAg>   Funding Agency/Sponsor
- make sure OntarioBuys and OCUL are listed
- put in links here as well (get from Appendix A)

    2.1.4.3   <depositr>   Depositor

Example 3:

<depositr abbr="" affiliation="**Gallup Canada Inc**.">**Gallup Canada Inc.**</depositr>
<depositr abbr="**CU**" affiliation="**Carleton University**">**Data Centre**</depositr>

    In this example, the Gallup polls were put into <odesi> by the Carleton University Data Centre.

Appendix F: Notes and Comments Tags – More Information

This appendix contains many different examples that could go in the "Notes and Comments" Tags.

- They are for any other information which you deem to be important to the study but which have not been mentioned elsewhere.
- These tags can be added to any section of your DDI document.

Check these examples out so you can get an idea of the information that you can put in there – feel free to use them!

These examples have been taken from the DDI pages at:
http://www.icpsr.umich.edu/DDI/dtd/version2-1-all.html

Examples:
<docDscr><verStmt><notes resp="**Jane Smith**">**Additional information on derived variables has been added to this marked-up version of the documentation**.</notes></verStmt></docDscr>

<docDscr><citation><notes resp="**Jane Smith**">**This citation was prepared by the archive based on information received from the markup authors.**</notes></citation></docDscr>

<docSrc><verStmt><notes resp="**Jane Smith**">**The source codebook was produced from original hardcopy materials using Optical Character Recognition (OCR)**.</notes><verStmt> </docSrc>

<docSrc><notes>**A machine-readable version of the source codebook was supplied by the Zentralarchiv**</notes></docSrc>

<docDscr><notes>**This Document Description, or header information, can be used within an electronic resource discovery environment**.</notes></docDscr>

<stdyDscr><verStmt><notes resp="**Jane Smith**">**Data for 1998 have been added to this version of the data collection.**</notes></verStmt></stdyDscr>

<stdyDscr><citation><notes resp="**Jane Smith"**>**This citation was sent to ICPSR by the agency depositing the data.**</notes></citation></stdyDscr>

<stdyInfo><notes>**Data on employment and income refer to the preceding year, although demographic data refer to the time of the survey.**</notes></stdyInfo>

<method><notes>**Undocumented codes were found in this data collection. Missing data are represented by blanks.</**notes></method>

<method><notes>**For this collection, which focuses on employment, unemployment, and gender equality, data from EUROBAROMETER 44.3: HEALTH CARE ISSUES AND PUBLIC SECURITY, FEBRUARY-APRIL 1996 (ICPSR 6752) were merged with an oversample.**</notes></method>

<setAvail><notes> **Data from the Bureau of Labor Statistics used in the analyses for the final report are not provided as part of this collection**.</notes></setAvail>

<dataAccs><notes>**Users should note that this is a beta version of the data. The investigators therefore request that users who encounter any problems with the dataset contact them at the above address**.</notes></dataAccs>

<fileStrc><notes>**The number of arrest records for an individual is dependent on the number of arrests an offender had**.</notes></fileStrc>

<fileTxt><verStmt><notes>**Data for all previously-embargoed variables are now available in this version of the file.**</notes></verStmt></fileTxt>

<fileDscr><notes>**There is a restricted version of this file containing confidential information, access to which is controlled by the principal investigator**.</notes></fileDscr>

<varGrp><notes>**This variable group was created for the purpose of combining all derived variables.**</notes></varGrp>

<varGrp><notes source="**archive**" resp="**John Data**">**This variable group and all other variable groups in this data file were organized according to a schema developed by the adhoc advisory committee.** </notes></varGrp>

<nCubeGrp><notes>**This nCube Group was created for the purpose of presenting a cross-tabulation between variables "Tenure" and "Age of householder."**</notes></nCubeGrp>

<valrng><notes subject="**political party**">**Starting with Euro-Barometer 2 the coding of this variable has been standardized following an approximate ordering of each country's political parties along a "left" to "right" continuum in the first digit of the codes. Parties coded 01-39 are generally considered on the "left", those coded 40-49 in the "center", and those coded 60-89 on the "right" of the political spectrum. Parties coded 50-59 cannot be readily located in the traditional meaning of "left" and "right". The second digit of the codes is not significant to the "left-right" ordering. Codes 90-99 contain the response "other party" and various missing data responses. Users may modify these codings or part of these codings in order to suit their specific needs.**</notes> </valrng>

<invalrng><notes>**Codes 90-99 contain the response "other party" and various missing data responses.** </notes></invalrng>

<var><verStmt><notes>**The labels for categories 01 and 02 for this variable, were inadvertently switched in the first version of this variable and have now been corrected.**</notes></verStmt></var>

<var><notes>**This variable was created by recoding location of residence to Census regions**.</notes></var>

<nCube><verStmt><notes>**The labels for categories 01 and 02 in dimension 1 were inadvertently switched in the first version of the cube, and have now been corrected.**</notes></verStmt></nCube>

<nCube><notes>**This nCube was created to meet the needs of local low income programs in determining eligibility for federal funds**.</notes></nCube>

<dataDscr><notes>**The variables in this study are identical to earlier waves.** </notes></dataDscr>

<otherMat><notes>**Users should be aware that this questionnaire was modified during the CAI process.**</notes></otherMat>