**Appendix A: Glossary**

**~~Add B2020 definition~~**

**B20/20**

* Beyond 20/20 is a software used to disseminate aggregate data. The B20/20 reader is freely downloadable and allows for some manipulation of aggregate data.
* http://www.beyond2020.com/

**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~~](http://www.ddialliance.org/codebook/index.html)~~)~~

The Data Documentation Initiative (DDI) is an international standard for describing statistical and social science data. Documenting data with DDI facilitates interpretation and understanding -- both by humans and computers. The freely available international DDI standard describes data that result from observational methods in the social, behavioral, economic, and health sciences. Use DDI to Document, Discover, and Interoperate!

<http://www.ddialliance.org/>

**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~~](http://www.ddialliance.org/DDI/related/tools.html)<http://www.ddialliance.org/resources/tools>

**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~~](http://www.ddialliance.org/codebook/index.html)~~)~~

Point to this page instead - <http://www.ddialliance.org/Specification/DDI-Codebook/2.5/>?

* 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.

OCUL

* Ontario Council of University Libraries
* ocul.on.ca

**~~ODESI~~ <odesi> \*\*change all reference to ODESI to <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~~ was initially funded by Ontario Buys and the OCUL Community. Currently, it is part of Scholars Portal and is supported by the OCUL Community and various other universities outside of Ontario on a subscription basis.

**~~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)~~

**OntarioBuys**

* Ontario Buys acted as the funded agency in the development of <odesi> from 2007 to 2009. The project was managed by OCUL and featured substantial start-up creation of DDI discovery metadata for research data in English as well as French.

**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 &amp; &lt; and &gt; 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**

* Scholars Portal is a service of the Ontario Council of University Libraries. Founded in 2002, Scholars Portal provides a shared technology infrastructure and shared collections for all 21 university libraries in the province.
* [~~http://www.scholarsportal.info/index.html~~](http://www.scholarsportal.info/index.html) <http://www.scholarsportal.info/>

**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.

1. 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

1. Note re: the data file:

When a .sav file is pulled into NESSTAR, it becomes an ~~NSDstat file~~ .Nesstar. Later on, if this file is exported into SPSS format, the new .sav file will be different from the original .sav file.

1. 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 please contact the ~~ODESI developers for more information at~~ ~~odesi@uoguelph.ca~~ team at Scholars Portal for more information.

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)~~ Statistics Canada Library (<http://www.statcan.gc.ca/eng/library/index>)
	+ 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

2.3.3.1   Response Rate

4.3.8.1 PreQuestion Text

4.3.8.2 Literal Question

4.3.8.3 PostQuestion Text

4.3.8.4 Forward Progression

4.3.8.5 Backflow

4.3.8.6 Interviewer Instructions

4.3.10 Range of Invalid Data Values

4.3.11 List of Undocumented Codes

4.3.18.1 Category Value

4.3.18.2 Label

4.3.18.4 Category Group Statistics

4.3.19 Coder Instructions

4.3.22.1 Derivation Description

4.3.22.2 Derivation Command

4.3.34 Variable Format

**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~~](http://www.ddialliance.org/codebook/index.html) <http://www.ddialliance.org/>

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 \*not sure whether or not we need this link.~~**

* ~~To keep in touch with new developments, post questions, …~~
* [~~http://odesi.uoguelph.ca/wiki/index.php/ODESI\_listserve~~](http://odesi.uoguelph.ca/wiki/index.php/ODESI_listserve) ~~is this current?~~

**~~Website~~**

* ~~odesi.ca~~
* [~~http://odesi.scholarsportal.info.proxy.library.carleton.ca/webview/~~](http://odesi.scholarsportal.info.proxy.library.carleton.ca/webview/)
* ~~This website is the gateway to find data, to learn more about <odesi>, related resources, and accessing data.~~
* ~~All metadata is downloadable. However, only Ontario Universities and other subscribing institutions are allowed to download data from this site. The only exception is for some data files that are open to the world.~~

**~~Wiki~~**

* ~~This wiki is open to all.~~
* [~~http://odesi.uoguelph.ca/wiki/index.php/Main\_Page~~](http://odesi.uoguelph.ca/wiki/index.php/Main_Page)
* ~~https://spotdocs.scholarsportal.info/display/odesi/introduction~~

**~~Tags specific to <odesi>~~**

**~~1.1.3.6   <fundAg> Funding Agency/Sponsor~~**

~~- make sure OntarioBuys and OCUL is are listed when appropriate~~

~~- 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~~*](http://www.icpsr.umich.edu/DDI/dtd/version2-1-all.html)~~is this current?~~

~~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>~~