WP 5
ENGLISH ONLY

UNITED NATIONS STATISTICAL COMMISSION and
ECONOMIC COMMISSION FOR EUROPE
CONFERENCE OF EUROPEAN STATISTICIANS

EUROPEAN COMMISSION
STATISTICAL OFFICE OF THE
EUROPEAN COMMUNITIES
(EUROSTAT)

ORGANISATION FOR ECONOMIC
COOPERATION AND DEVELOPMENT
(OECD)
STATISTICS DIRECTORATE

Joint UNECE/Eurostat/OECD work session on statistical metadata (METIS)
(Geneva, 9-11 February 2004)

Topic (ii): Metadata interchange

From Data to Knowledge: Standards and Systems for Sharing Data and Information across Organizations

Invited Paper

Submitted by Health Canada, Canada[1]

I INTRODUCTION

1 The focus is on metadata standards and tools for delivery of evidence-based knowledge to support policy development, program management and accountability in government. The emphasis is on standards for end-user data and information products, and especially for microdata and associated aggregates. The paper discusses the Data Documentation Initiative (DDI) standard and tools, methods that have been developing from the 1960s in university data archives and libraries to serve research and teaching needs, and describes how they have been adapted and extended in the DAIS|nesstar system for users and producers of official statistics in government. It demonstrates metadata management and dissemination tools for finding, exchanging, accessing, combining, and analyzing data and information across organizations on the Web. The need for metadata standardization at the data item level is emphasized, as is linking with associated information and knowledge products that standardization at the data item level enables. The DDI XML DTD is demonstrated as a workable standard that is available now for improved pay off and delivery from national and international statistical systems. The importance for return on investment in national statistical offices (NSOs) is discussed. Linkages to ISO 11179 and its Corporate Metadata Repository (CMR) statistical extension in the form of semantic mappings and conceptual models are introduced, as are relationships to the SDMX initiative.

2 In keeping with the work program of the meeting, the paper focuses on the following substantive topics, with emphasis on interchange:

(i) Functions of metadata in statistical production;
(ii) Metadata interchange;

[1] Prepared by Bill Bradley, Director, Data Systems and Standards Division, Applied Research and Analysis Directorate, Information, Analysis and Connectivity Branch Bill_Bradley@hc-sc.gc.ca tel. + 1 613 957-0702

(iii) Metadata models and terminology;
(iv) Use of metadata for searching and finding statistical data in websites and portals.

II The DDI - Statistical metadata for end-user research and analysis

3 The Data Documentation Initiative (DDI) was discussed in detail for METIS by Ryssevik in 1999. A description of the standard, its history, results of recent meetings, and the new DDI Alliance is available at http://www.icpsr.umich.edu/DDI/. The following provides a brief summary, update and analysis.

4 **Summary** The DDI is a recent standard for statistical metadata formulated by the Data Documentation Initiative, a committee of experts from the international data archive and library community and, at various times, several statistical agencies. Expressed as an XML DTD, it represents machine-readable data documentation (metadata) requirements as seen by data archivists and managers, research support staff and analysts who provide data services in universities and to some extent government. It has emerged from this data community's longstanding concerns for documentation to support effective access, use, management and preservation of statistical data resources for research and teaching.

5 **Background** The DDI was conceived during the meetings of the International Association for Social Science Information Service and Technology (IASSIST) which took place at the Universities of Edinburgh in 1993 and California at Berkeley in 1994, and initiated under the leadership of the Inter University Consortium for Political and Social Research (ICPSR) following IASSIST 1995 in Quebec City. Version 1, which provides a standard for microdata was developed by means of a series of meetings, including a full beta test, which took place from 1995 to 2000 through the support of ICPSR, with grants from the National Science Foundation (NSF). When the NSF grants ran out, bridge funding was provided by Health Canada's DAIS|nesstar project, making possible the development of Version 2 which was released in July 2003; it incorporates extensions for aggregate data and key elements of geography.

6 **DDI Alliance and Current Work Program** The DDI Committee was recently replaced by the Alliance for the DDI, whose members join in on a fee basis to continue the refinement, development and promulgation of the standard. Its first meeting took place at the University of Michigan in October 2003, and was attended by 30 stakeholders. The ensuing work program includes further refinements to the Version 2 extensions for aggregate tabular data and geography, the development of XML schema and conceptual data model representations, harmonization with ISO 11179 and its statistical extensions, outreach to new areas especially data producers and other standards developers, and establishment of a stable funding base. As the DDI's standards for micro and aggregate data mature to cover time series data, there is an interest in harmonization with SDMX.

7 **Evaluation of the DDI** Since its first release in March 2000, the DDI XML DTD has achieved rapid take up amongst the data library and archive community as a means for exchanging metadata that describes survey, census, administrative microdata, and more recently aggregate data from official statistics organizations. At the request of the National Science Foundation, ICPSR conducted an external evaluation of the DDI over the period 1997-2000. The evaluators were Daniel Greenstein, Director, Digital Libraries Federation; James Jacobs, Data Services Librarian, University of California, San Diego; Tom W. Smith, Director, U. S. General Social Survey and Stuart Weibel, Director, Dublin Core Metadata Initiative.

8 'The evaluators were struck by how quickly the activity had moved forward in comparison with other standards-setting projects, and were impressed by the enthusiastic buy-in that has taken place and the investments of the many stakeholders in the DDI.' (NSF Report, 2001) 'Perhaps the most important impact of ... the DDI effort is that the data will be far easier to process, analyse and compare, greatly increasing the impact of the funds spent on creating the data initially, and creating new opportunities to use the data productively. Without efforts such as this, such use is limited by the costly effort of hand-crafted analysis.' (Weibel, 2001) The success of the DDI can be attributed in part to the following factors.

- **Results-based approach** The DDI was conceived in practical terms as a simple, purpose-based standard, thereby enabling its basic elements to be formulated and implemented rapidly. The goal was to implement a modern, machine-readable data documentation and exchange standard for use by data producers, managers, and research organizations, and especially to eliminate the enormous duplication of effort which must take place for common data sources to be used effectively in different processing and support environments.

- **Metadata as a means to an end - data dissemination, access and use** The DDI grew from a strong tradition of data sharing and exchange for secondary analysis. It was influenced by and benefited from the OSIRIS codebook, itself a de facto standard for machine readable data documentation that had been used for disseminating data to libraries and archives world wide by ICPSR from the mid 1960s; the efforts of Roistacher (1978) to define a style manual for machine readable data dictionaries; the generic data dictionary and codebook implemented in a simple relational structure in the Data Dictionary Management System (DDMS) by Health Canada in the 1980s (Bradley et al, 1991); and the CASES computer assisted interviewing system and its companion Survey Data Analysis (SDA) package from UC Berkeley. The result was a generic data dictionary composed of descriptors which can be used to generate programs in the languages of most statistical data processing packages, enhanced with elements at both the data item and data set/study description level that are needed by people to make sensible use of the described data elements.
- **Built on previous standards** In a very basic sense, the goal was to move previous de facto standards, such as the OSIRIS unit record machine readable codebook and the DDMS pseudo relational DBF structure, forward into the world of modern text-oriented data base schema. The objective was to create a document type definition (DTD) in accordance with the ISO Standard Generalized Markup Language (SGML).
- **Focus on semantic structures first, data models later** The Committee therefore focused on the elements of data documentation to be included in the standard, their meaning, and their logical structures, data types and inter-relationships in the form of a DTD representation. No thought was given to a data model representation - indeed the view was that a data model such as that embodied in the DDMS system was outmoded, and inappropriate for metadata which, after all, is primarily text. In fact a portion of the new standard, as well as the Committee's first SGML programmers, was borrowed from the Text Encoding Initiative (TEI), a humanities computing undertaking to mark up key works of literature (http://helmer.aksis.uib.no/humdata/hd90-3/hd390lou.htm). As a result the DDI has been largely intuitive to most data professionals, many of whom were soon marking up data documentation directly, without the benefit of specialized software tools or programmer intervention. A data model would not have been intuitive for those whose job it is to populate the metabase.
- **Sensible approach to the data food chain** At its first meeting in Quebec City, 1995, the Committee decided to begin by focusing on the key metadata requirements for survey microdata, leaving aside aggregate and time series data, which would be addressed later. Since time series are often constructed from aggregates, and aggregates in turn are usually based on observations at microdata level, it seemed to make sense to build the standard from the bottom up - that is to start by focusing on the metadata for what had been counted in the first place, before dealing with the counts themselves or with counts or other standard aggregates that occur over time with a common periodicity. Indeed, some argued that to start standardization efforts at the macro data level would be to risk making errors of specification, since the underlying requirements for metadata at the micro data level might not have been sufficiently formulated.
- **Focus on plumbing, not content** The focus of the DDI was entirely on the descriptors to be used, and their structure (the 'plumbing'); there was no concern with standardizing the content that would flow through the pipes, nor how that content should be classified. The latter are far more difficult tasks, perhaps impossible given the range of data resources and producers to which the work was intended to apply. The strategy was to make the metadata observable in a standardized way first - the standardization of concepts, content and ways to classify it would emerge later in ways that made sense from the observed content for the application at hand.
- **Staged approach in manageable steps** Neither was the Committee concerned with perfection. More difficult issues, such as methods for handling complex files, or how to deal with the complicated skip logic structures that can occur in some computer assisted surveys, were, like the problem of describing aggregate and time-series data, set aside for later. The first goal was to handle simple micro datasets in flat file format, since the most statistical datasets are in this form.
- **User visibility, testing and feedback** The result was that by the 1996 meeting of IASSIST - one year after the initiative began - the core of the standard, expressed as an SGML DTD, was in place. The

Committee was able to present the DTD to professional groups and potential users, who were able to try it out using standard SGML tools and provide feedback. Moreover it was a simple matter to convert the DTD to XML almost immediately after the initial formulation of XML by W3C in 1997, thereby increasing the momentum amongst the user community. Further honing was carried out; a tag library manual prepared, and a comprehensive beta test was conducted by many organizations in North America and Europe.

III Development of DDI support and exploitation tools

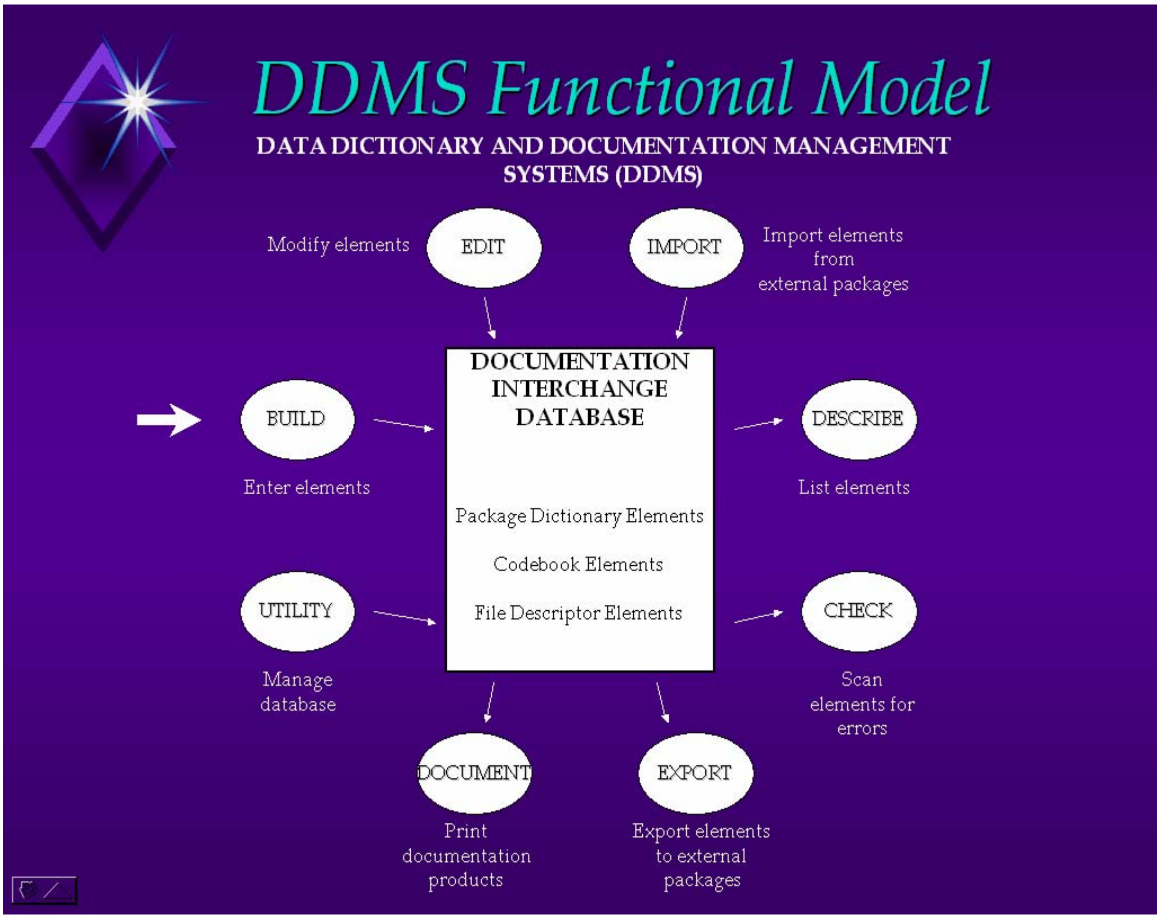
9 Standards are of little practical value without tools and systems to support and exploit them. The success of the DDI is in no small part due to the fact that it emerged from a strong, tool-based environment that had already engendered a base of easily adaptable machine-readable metadata. In addition, modern web-based tools emerged quickly.

10 Tools that preceded the DDI were the codebook creation and management elements of the OSIRIS data processing system developed by ICPSR beginning in the late 1960s, and still used by ICPSR to transfer data documentation to its members; and the Data Documentation Management System (DDMS) and its companion Data and Information Sharing system (DAIS) used by Health Canada from the mid 1980s, to provide access to data and information resources through standardized metadata. The key tool that followed the DDI was the Nesstar system, and more recently the DAIS|nesstar suite that has emerged as a result of collaboration between Nesstar Ltd. and Health Canada. The following paragraphs summarize.

11 **OSIRIS** In use from 1967, ICPSR's **O**rganized **S**et of **I**ntegrated **R**outines for **I**nvestigation with **S**tatistics (Barje and Marks, 1973) was the world's first integrated statistical package (ISP), whereby different statistical routines operate from a common data dictionary and dataset structure and make use of common data management, manipulation and transformation functions. OSIRIS' metadata, however, went well beyond the requirements for machine processing to include elements needed by users to exchange and make intelligent secondary use of data. OSIRIS codebook elements included file descriptors and unit-record characteristics, question text, interpretative notes and information about the provenance and methodology associated with the data.

12 Although the statistical aspects of the OSIRIS system were superseded for most users by SPSS and SAS in the 1970s, both of which made use of OSIRIS data dictionary elements in their integrated metadata components, the OSIRIS Codebook continued to be used by ICPSR to transfer machine-readable data documentation to its member institutions. As such the OSIRIS Codebook became a de facto though increasingly outmoded standard for metadata interchange within this community. Early versions of SPSS included a facility for reading OSIRIS codebooks, and efforts were made to define and promote an OSIRIS Codebook-based data interchange file that would be read and written by the growing number of ISPs (Roistacher, 1976).

13 **DDMS/DAIS** Development of the DDMS/DAIS model and system (Bradley et al, 1994) began in the 1980s during the early stages of the micro-computer revolution as a means for creating and managing a standardized database of metadata through which data resources could be exchanged and accessed via networks at the desktop. The work was strongly influenced by the OSIRIS standard and the data interchange efforts of Roistacher (1976), but adapted these to the data and information requirements of users of official statistics data in a large government organization.



DDMS Functional Model

DATA DICTIONARY AND DOCUMENTATION MANAGEMENT SYSTEMS (DDMS)

Modify elements EDIT

IMPORT Import elements from external packages

DOCUMENTATION INTERCHANGE DATABASE

Package Dictionary Elements

Codebook Elements

File Descriptor Elements

BUILD Enter elements

DESCRIBE List elements

UTILITY Manage database

CHECK Scan elements for errors

DOCUMENT Print documentation products

EXPORT Export elements to external packages

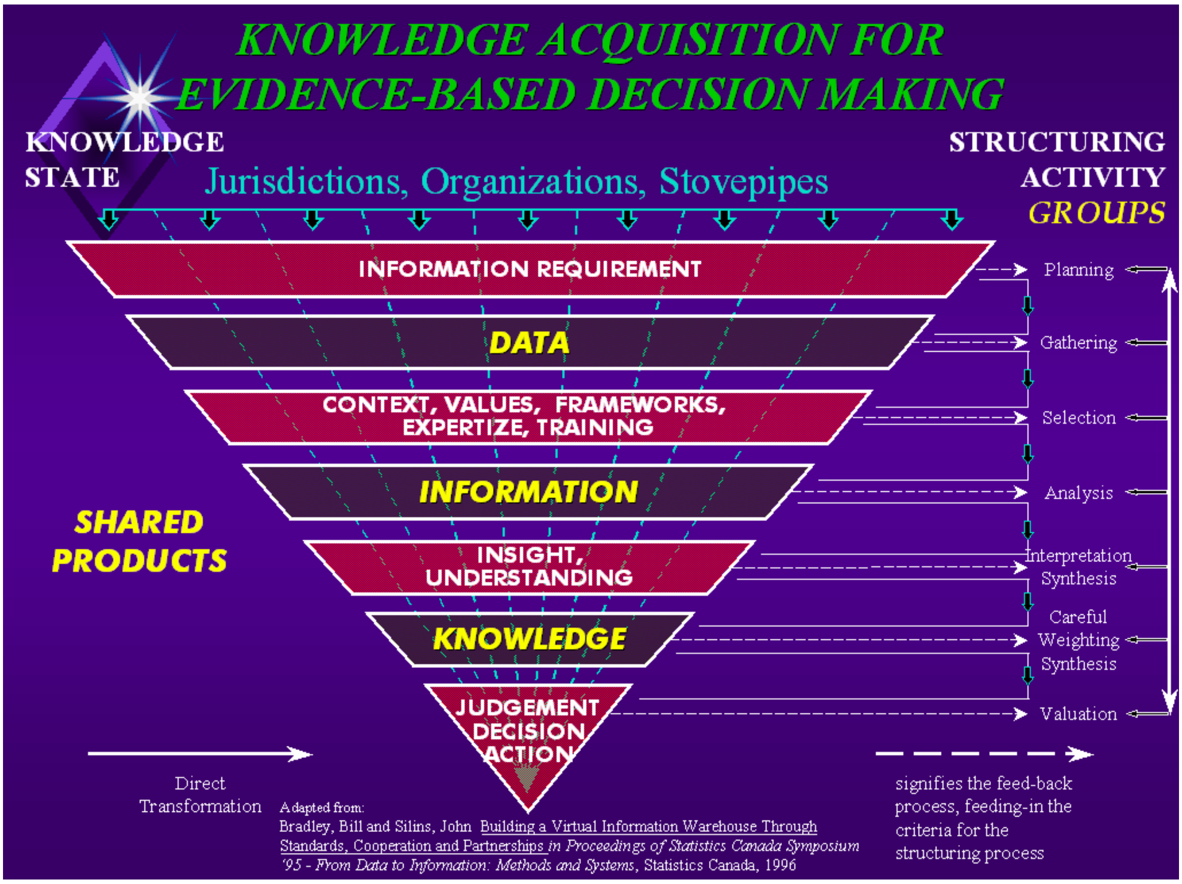
Figure 1 Original functional model for DDMS system, 1986

14 First to be created was The Data Documentation Management System (DDMS) (Bradley et al, 1990), a PC-based tool for creating, managing and transporting data dictionaries and documentation between statistical packages and users. Although the software implementation eventually became outdated, in the view of some the functional model for DDMS, depicted in Figure 1, ‘remains valid today’ (Bargmeyer, 2000). The first version of DDMS went into production in 1986, and the system has been used to document all health and social policy microdata of interest to Health Canada from the late 1970s through 2003, when the DdiMS software discussed below replaced it.

15 One of the purposes of DDMS was to reduce duplication of effort in metadata creation by enabling metadata to be exchanged and shared, and ideally to be created once and properly at source for dissemination with the data by data producers. To these ends DDMS was used by the Canadian Association of Research Libraries to create shareable metadata for Canadian census data, and by the federal-provincial Canadian Heart Health Initiative to describe and integrate provincial surveys. It was used also at various times by Health Canada’s key data suppliers, including the Census of Canada, Statistics Canada’s General Social Survey and Post Censal Survey programs, and private sector services such as the Canada Health Monitor.

16 As the standardized DDMS metabase grew to include more surveys, and as the local area networking technology in Health Canada matured to link more desktops, planning and promotion of both metadata standardization and a client server application to provide desktop access to the Department’s data treasures proceeded. A prototype Data and Information Sharing (DAIS) client was prepared for demonstration purposes, and the DDMS/DAIS model articulated (Bradley et al, 1994 part 2). The prototype included a demonstration of what proved to be the key requirement for obtaining support for metadata standards activities in this environment - a focus on the role of metadata in the acquisition, management and delivery of knowledge for policy decision-making. We learned that metadata standardization is just as important for knowledge delivery and payoff in policy environments as it is for statistical production or exchange in NSOs or international agencies.

17 The DDMS/DAIS system is therefore based on the 'premise that the goal of statistical activities is the creation and exploitation of evidence-based knowledge. That is, that statistical data and information are of little value until they result in knowledge, and, indeed, until that knowledge contributes to a decision or action that adds value or recognized improvement in the domain of application.' (Bradley et al, 2003) In order to maximise payoffs from expenditures on statistical information, statistical agencies, and especially metadata specialists, need to understand how statistical data are transformed into information and knowledge, and adopt practices that facilitate such transformations.



Adapted from:
Bradley, Bill and Silins, John Building a Virtual Information Warehouse Through Standards, Cooperation and Partnerships *in Proceedings of Statistics Canada Symposium '95 - From Data to Information: Methods and Systems*, Statistics Canada, 1996

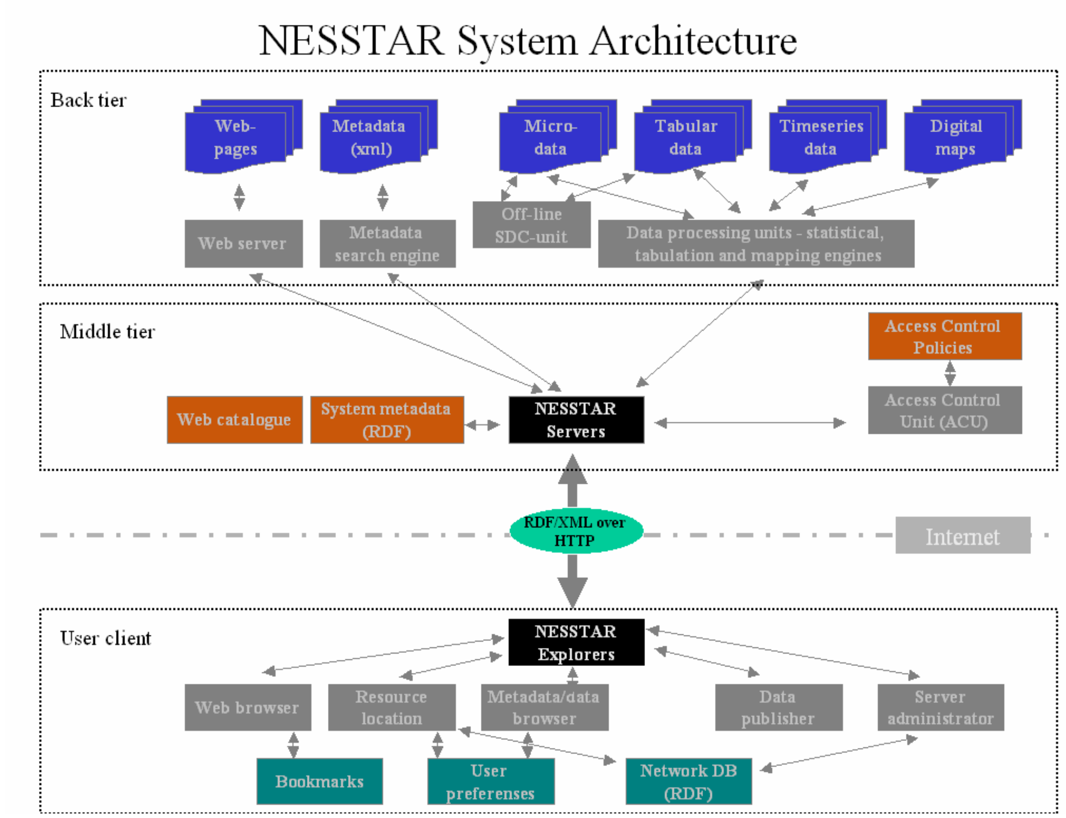
Figure 2 DAIS knowledge acquisition model

18 The knowledge acquisition model, Figure 2, is a discussion tool for helping to develop and refine such an understanding, and for guiding our thinking about how DAIS should function. It was devised by Alexander (1990), then extended and elaborated by Bradley et al (1994), and Bradley and Silins (1995). The inverted pyramid depicts the vast store of data resources at the base, and the selection and filtering processes that take place as data are converted to information, and information to knowledge. Knowledge is a highly individuated commodity - so also the supporting evidence, *which is why custom analyses from microdata are so important*. Workers need to have easy access to the pyramid from any level, depending upon their requirements, orientation and skills. No matter the level of entry, more often than not the data and information resources needed are in the custody of other organizations, and when access is achieved the analysis and transformation processes are highly iterative, requiring 'drill-down' to data, or 'drill-up' to information or knowledge. The key functional requirements are therefore the ability to move in all directions in the pyramid as rapidly as possible, and especially to be able to do custom analyses from micro data, fully linked with previous value added in the form of information (tables, analyses, time-series, indicators) and knowledge (reports, fact sheets, publications).

19 **NESSTAR** The rapid formulation of an initial DTD by the DDI Committee enabled a group consisting of the Danish Data Archive, the UK Data Archive, and the Norwegian Social Science Data Service (NSD) to obtain funding to create supporting software tools and services. Building on the integrated data catalogue activity of the Council of European Social Science Data Archives (CESSDA) in 1995, and

under the auspices of the Telematics Applications Programme, an activity under DGXIII of the European Commission's 4th Framework Programme, the **N**e**t**worked **S**ocial **S**cience **T**ools **a**nd **R**esources (NESSTAR) Project came into being for a two year period in January 1998. The result was that by the time Version 1 of the DDI was ready for release in 2000, there was already a set of tools to support the creation of DDI XML instances and demonstrate the power of such standardization in a distributed server environment across the web. Nesstar's funding was extended for a further two year period to move the work forward from micro data into the realms of aggregate and time-series data.

20 Nesstar's system architecture, as depicted in a presentation by Bradley, Musgrave and Ryssevik in 2001, is shown below. The distributed server technology included **1)** an XML search engine licensed from a California university by means of which DDI descriptors at the study (dataset) level could be searched across the web **2)** a browsing capability at the dataset and variable level whereby the detailed contents of each dataset could then be ascertained via its DDI descriptors **3)** an access control unit (ACU) through which the server administrator could control access via password control depending on the permissions granted to the user by the local data authority **4)** a web version of NSD's powerful NSDStat statistical engine, by means of which simple tabular, regression and other analyses could be performed and visualized across the web from data housed on remote servers by those having permission to do so **5)** NSDStat's capabilities for exporting data in various popular statistical formats, by means of which selected subsets of data could be downloaded across the web to the local workstation for detailed analysis, again by those having the appropriate permission from the originating data authority.



21 Included also were various utilities for **1)** capturing the available metadata from existing data sets in SPSS and NSDStat format **2)** adding additional DDI fields such as question text at the variable level and study descriptors at the data set level **3)** creating and editing DDI XML instances **4)** adding DDI-specified descriptive statistics via the NSD statistical engine, and **5)** 'publishing' the metadata and data to Nesstar servers. The client software consisted of a large java application (the Nesstar Explorer) and a light, server-based version operable via a java-enabled web browser. And all of this impressively demonstrable across several data archive sites on the web by the time the first version of the DDI DTD was officially released in March 2000!

22 **DAIS|nesstar** By March 2000 the DAIS client was installed on several hundred desktops across Health Canada, and providing access to the thousands of data items and questions that had resulted from the more than $1B in data investments in the health and social policy domain in Canada since the 1970s. Reflecting the knowledge acquisition model displayed in Fig. 2, the data were fully linked through their standardized metadata with associated tables, reports and indicators. The underlying systems, metadata and standardization work was by this time being supported by the Department's new Applied Research and Analysis Directorate, as a key element of a program to improve accountability and the use of evidence throughout the Department, as well as the health system as a whole. There was now strong support for metadata standards and standardization, and funding was available to prepare a more modern, sustainable version of the software, one that exploited the capabilities of the web.

23 With support available to move the DAIS application to the web, it made sense that the DDMS/DAIS and Nesstar projects should join forces. There had already been considerable cross-fertilisation. The prototype versions of DDMS and DAIS, along with associated metadata standardization and creation issues, had been the subject of an extended visit by the author to the UK Data Archive in 1993, when much of the 1994 'Metadata Matters' papers had been drafted, and the notion of an SGML implementation articulated (Musgrave, 1993). Collaboration of a professional nature had continued, especially in the context of the ensuing standards development activities, and the latest version of DAIS in Health Canada was benefiting from the use of the NSD statistical engine.

24 Moreover, the new Nesstar system very closely resembled DDMS/DAIS and already provided DAIS' basic functions in the web environment. Nesstar's distributed web server technology was fully consistent with the DAIS vision (Bradley et al 1994), which, it was believed, would be ideal for serving the distributed data and information sharing requirements of large government organizations in a multi-jurisdictional federal state. Given the desirability of facilitating evidence-based decision-making throughout the whole health system, where there are many knowledge-intensive actors at all levels from local and regional through international, the existing distributed server technology seemed to provide an excellent starting point.

25 The difference was that DAIS' capabilities were much more extensive in ways that had been proven to be necessary for use in the applied research, official statistics, program management and policy milieu of Health Canada. Where Nesstar had been designed mostly with the needs of the academic community in mind, DAIS brought the requirements and methods of this official statistics and policy community to the table.

26 Health Canada had also foreseen the need for an independent, self-supporting infrastructure to support the tools and help to govern data and information access, operational standardization and web services across servers and organizations. (Bradley et al, 1994), and the Nesstar project had in mind the possibility of becoming a self-supporting commercial entity. Bearing in mind as well the historically fragile nature of metadata standards endeavours and the investment in tools and infrastructure required to make such standardization a reality, it made sense from a governance perspective not to facilitate the development of competing solutions, but rather to strengthen and build on what was there already.

27 The DAIS|nesstar project was initiated in January 2001. A demonstration version using the existing Nesstar server was implemented during the first three months, and then both client and server were re-engineered to incorporate the DAIS modifications needed to make the tools tractable in the environment of a large government department. These included:

- Development of DdiMS, the fully integrated metadata management tool for the DDI based on the DDMS functional model and fifteen years of operational metadata standardization experience and refinements
- Elimination of statistics that are usually invalid for official statistical data
- Addition of security (secure socket layer encoding) equivalent to electronic banking
- Addition of information ('tables') and knowledge ('fact sheets') products as shared objects in the system, as outlined in Figure 2. Metadata for these are based on the Dublin Core
- Elimination of long, unstructured browse lists, which were replaced with DAIS' telescoping lists
- Addition of a shared 'group' object for structuring purposes (Figure 2), such that results from browsing or searching can be refined, saved and retrieved for future use

- Customizable four panel display to help users simplify the multiple-windows situations that typically arise in knowledge development applications involving data and information resources from multiple sources
- Ability to search for any or all objects – fact sheets, tables, data sets, variables – and to save and refine search results for further use as one or more ‘groups’ or private lists
- Re-engineering of server to incorporate a relational database structure and engine, needed to effect most DAIS extensions
- Ability to ‘drill-down’ through the pyramid, from knowledge products to underlying aggregates and micro data, to permit fast re-analysis from the perspective of different assumptions, definitions or needs
- Ability to ‘drill-up’ from a data item or dataset to the tables, reports and publications produced from it
- Extended access control capabilities, for example to permit viewing of metadata for proprietary research reports while restricting access to the reports themselves, or restricting access to proprietary metadata fields such as question text and frequencies for private sector surveys
- Generation, graphical exploration, saving and retrieval of multidimensional cubes according to the DDI aggregate data standard
- Ability to access published tables in Excel, PDF or Beyond 2020 web server or browser format. In addition to DDI cubes, other publication formats, such as Supercross or PC-Axis can be added as needed
- A ‘compare’ tool to assist in building groups of comparable variables across surveys based on an examination of key metadata fields, with capabilities for recoding and selecting subsets of cases as needed
- A wizard for displaying univariate trends or time series based on comparable variables across different surveys over time, or to carry out comparative analyses for different subpopulations or geographies
- Automated tabulation of trends or time series using comparable independent variables across different data sets, or across subpopulations or geographies, with visual exploration of resulting cubes or maps

28 The following sections introduce and demonstrate DDI and DAIS|nesstar capabilities which pertain to the meeting’s work program.

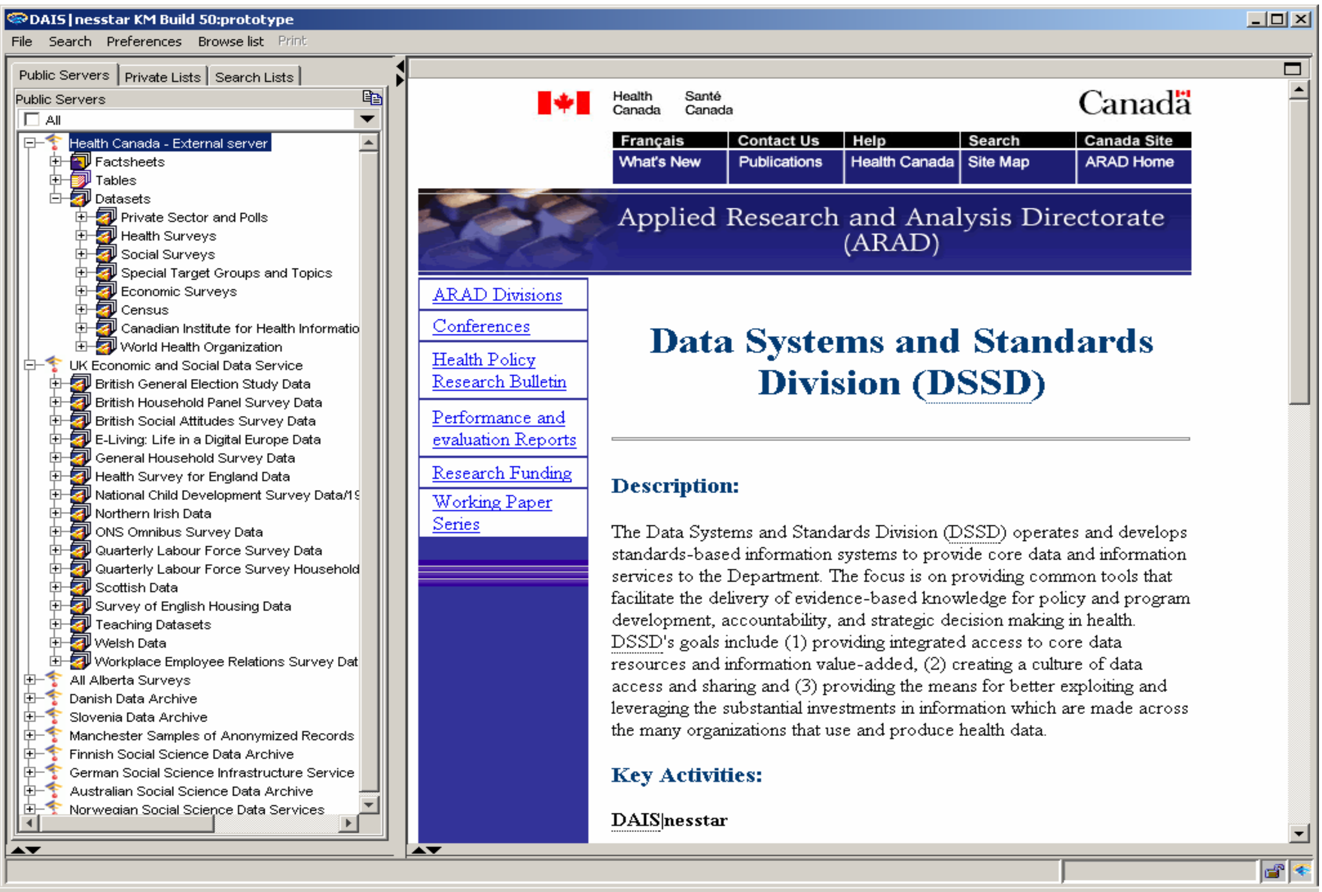
IV Metadata Interchange

29 Figure 9, Section V, shows a small portion of DAIS|nesstar’s DdiMS process for creating an XML-tagged documentation interchange database file (Figure 1) as a text database that is compliant with the DDI DTD. The XML file can then be exchanged physically for direct manipulation as text by knowledgeable users, or for processing by software programmed to read XML structured according to this DTD, such as DdiMS and DAIS|nesstar clients. Figures 4, 5 and 6 provide views of the documentation contained in DDI instances, as provided by the DAIS|nesstar Knowledge Maker (KM) client for browsing purposes.

30 The left-hand panel of Figure 4 shows some of the demonstration servers that are presently accessible through DAIS|nesstar clients. The Health Canada and UK Economic and Social Data Service servers have been expanded to show their top-level browse lists, while the right hand pane shows the web home page for the server highlighted.

31 Figure 5 shows browsing at the data set level. The DDI summary descriptors for the highlighted dataset (or what academic data librarians call ‘study description’) are available through tabs in the lower right panel, as prepared by DDI-knowledgeable server support staff. The detailed documentation provided with the data set by the data supplier can be expanded for reference purposes through the upper right panel.

32 In Figure 6 the codebook containing the data item descriptions for this survey has been expanded, and the metadata for the question on self reported health status is being examined. A summary table and chart showing question text and useful univariate statistics forms the default display, and other DDI descriptors are available through the various tabs.


DAIS|nesstar KM Build 50:prototype
File
Search
Preferences
Browse list
Print
Public Servers
Private Lists
Search Lists
Public Servers
All
Health Canada - External server
Factsheets
Tables
Datasets
Private Sector and Polls
Health Surveys
Social Surveys
Special Target Groups and Topics
Economic Surveys
Census
Canadian Institute for Health Informatio
World Health Organization
UK Economic and Social Data Service
British General Election Study Data
British Household Panel Survey Data
British Social Attitudes Survey Data
E-Living: Life in a Digital Europe Data
General Household Survey Data
Health Survey for England Data
National Child Development Survey Data/19
Northern Irish Data
ONS Omnibus Survey Data
Quarterly Labour Force Survey Data
Quarterly Labour Force Survey Household
Scottish Data
Survey of English Housing Data
Teaching Datasets
Welsh Data
Workplace Employee Relations Survey Dat
All Alberta Surveys
Danish Data Archive
Slovenia Data Archive
Manchester Samples of Anonymized Records
Finnish Social Science Data Archive
German Social Science Infrastructure Service
Australian Social Science Data Archive
Norwegian Social Science Data Services
Health Canada
Santé Canada
Canadä
Français
Contact Us
Help
Search
Canada Site
What's New
Publications
Health Canada
Site Map
ARAD Home
Applied Research and Analysis Directorate (ARAD)
ARAD Divisions
Conferences
Health Policy Research Bulletin
Performance and evaluation Reports
Research Funding
Working Paper Series

Data Systems and Standards Division (DSSD)

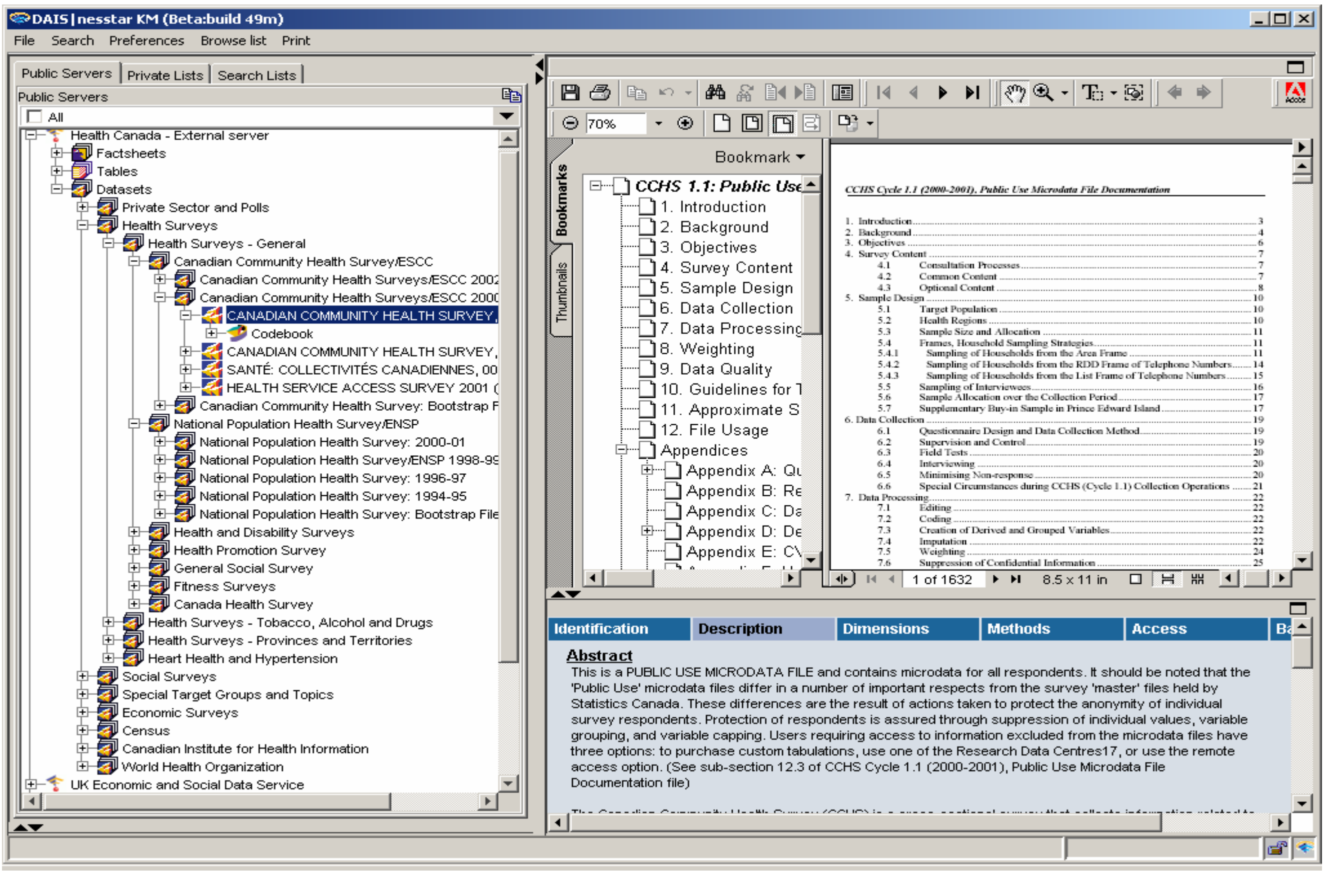
Description:

The Data Systems and Standards Division (DSSD) operates and develops standards-based information systems to provide core data and information services to the Department. The focus is on providing common tools that facilitate the delivery of evidence-based knowledge for policy and program development, accountability, and strategic decision making in health. DSSD's goals include (1) providing integrated access to core data resources and information value-added, (2) creating a culture of data access and sharing and (3) providing the means for better exploiting and leveraging the substantial investments in information which are made across the many organizations that use and produce health data.

Key Activities:

DAIS|nesstar

Figure 4 Present demonstration servers are listed on left. Health Canada and UK browse lists are expanding

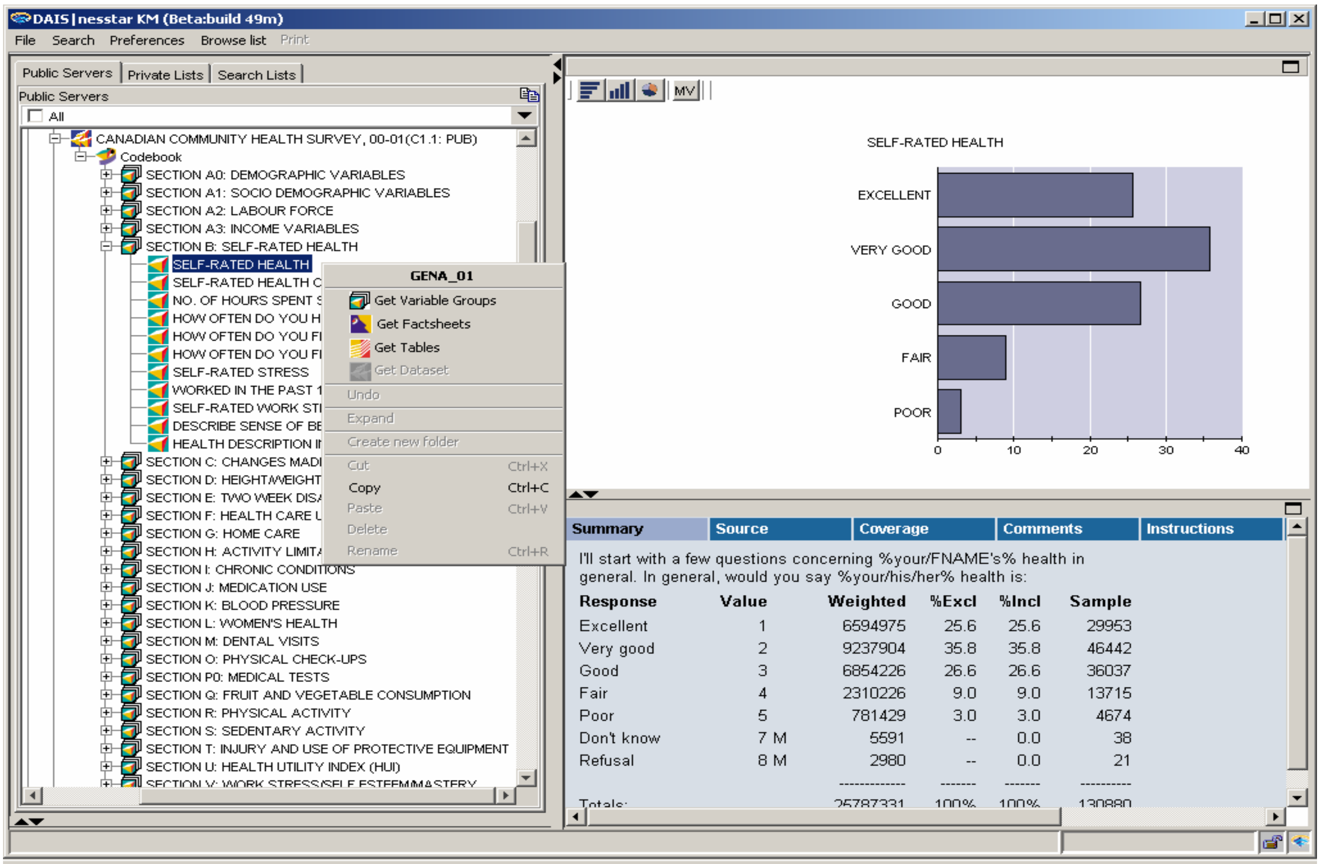

DAIS|nesstar KM (Beta:build 49m)
File
Search
Preferences
Browse list
Print
Public Servers
Private Lists
Search Lists
Public Servers
All
Health Canada - External server
Factsheets
Tables
Datasets
Private Sector and Polls
Health Surveys
Health Surveys - General
Canadian Community Health Survey/ESCC
Canadian Community Health Surveys/ESCC 2002
Canadian Community Health Surveys/ESCC 2000
CANADIAN COMMUNITY HEALTH SURVEY,
Codebook
CANADIAN COMMUNITY HEALTH SURVEY,
SANTÉ: COLLECTIVITÉS CANADIENNES, 00
HEALTH SERVICE ACCESS SURVEY 2001 (
Canadian Community Health Survey: Bootstrap F
National Population Health Survey/ENSP
National Population Health Survey: 2000-01
National Population Health Survey/ENSP 1998-99
National Population Health Survey: 1996-97
National Population Health Survey: 1994-95
National Population Health Survey: Bootstrap File
Health and Disability Surveys
Health Promotion Survey
General Social Survey
Fitness Surveys
Canada Health Survey
Health Surveys - Tobacco, Alcohol and Drugs
Health Surveys - Provinces and Territories
Heart Health and Hypertension
Social Surveys
Special Target Groups and Topics
Economic Surveys
Census
Canadian Institute for Health Information
World Health Organization
UK Economic and Social Data Service
70%
Bookmarks
Thumbnails
Bookmark
CCHS 1.1: Public Use
1. Introduction
2. Background
3. Objectives
4. Survey Content
5. Sample Design
6. Data Collection
7. Data Processing
8. Weighting
9. Data Quality
10. Guidelines for T
11. Approximate S
12. File Usage
Appendices
Appendix A: Qu
Appendix B: Re
Appendix C: Da
Appendix D: De
Appendix E: C\
1 of 1632
8.5 x 11 in
Identification
Description
Dimensions
Methods
Access
Ba

CCHS Cycle 1.1 (2000-2001), Public Use Microdata File Documentation

Abstract

This is a PUBLIC USE MICRODATA FILE and contains microdata for all respondents. It should be noted that the 'Public Use' microdata files differ in a number of important respects from the survey 'master' files held by Statistics Canada. These differences are the result of actions taken to protect the anonymity of individual survey respondents. Protection of respondents is assured through suppression of individual values, variable grouping, and variable capping. Users requiring access to information excluded from the microdata files have three options: to purchase custom tabulations, use one of the Research Data Centres17, or use the remote access option. (See sub-section 12.3 of CCHS Cycle 1.1 (2000-2001), Public Use Microdata File Documentation file)

Figure 5 Browsing metadata at the data set level. DDI descriptors are viewed through summary tabs in lower right panel, while detailed source documentation is in the upper panel



CANADIAN COMMUNITY HEALTH SURVEY, 00-01(C1.1: PUB)
Codebook
SECTION A0: DEMOGRAPHIC VARIABLES
SECTION A1: SOCIO DEMOGRAPHIC VARIABLES
SECTION A2: LABOUR FORCE
SECTION A3: INCOME VARIABLES
SECTION B: SELF-RATED HEALTH
SELF-RATED HEALTH
SELF-RATED HEALTH C
NO. OF HOURS SPENT S
HOW OFTEN DO YOU H
HOW OFTEN DO YOU FI
HOW OFTEN DO YOU FI
SELF-RATED STRESS
WORKED IN THE PAST 1
SELF-RATED WORK STI
DESCRIBE SENSE OF BE
HEALTH DESCRIPTION II
SECTION C: CHANGES MADI
SECTION D: HEIGHT/WEIGHT
SECTION E: TWO WEEK DISA
SECTION F: HEALTH CARE U
SECTION G: HOME CARE
SECTION H: ACTIVITY LIMITA
SECTION I: CHRONIC CONDITIONS
SECTION J: MEDICATION USE
SECTION K: BLOOD PRESSURE
SECTION L: WOMEN'S HEALTH
SECTION M: DENTAL VISITS
SECTION O: PHYSICAL CHECK-UPS
SECTION P0: MEDICAL TESTS
SECTION Q: FRUIT AND VEGETABLE CONSUMPTION
SECTION R: PHYSICAL ACTIVITY
SECTION S: SEDENTARY ACTIVITY
SECTION T: INJURY AND USE OF PROTECTIVE EQUIPMENT
SECTION U: HEALTH UTILITY INDEX (HUI)
SECTION V: WORK STRESS/SELF ESTEEM/MASTERY

I'll start with a few questions concerning %your/FNAME's% health in general. In general, would you say %your/his/her% health is:

Response	Value	Weighted	%Excl	%Incl	Sample
Excellent	1	6594975	25.6	25.6	29953
Very good	2	9237904	35.8	35.8	46442
Good	3	6854226	26.6	26.6	36037
Fair	4	2310226	9.0	9.0	13715
Poor	5	781429	3.0	3.0	4674
Don't know	7 M	5591	--	0.0	38
Refusal	8 M	2980	--	0.0	21
		------------	-------	-------	----------
Totals:		25787331	100%	100%	130880

Figure 6 Metadata browsing at the data item (variable) level

V Functions of Metadata in Statistical Production

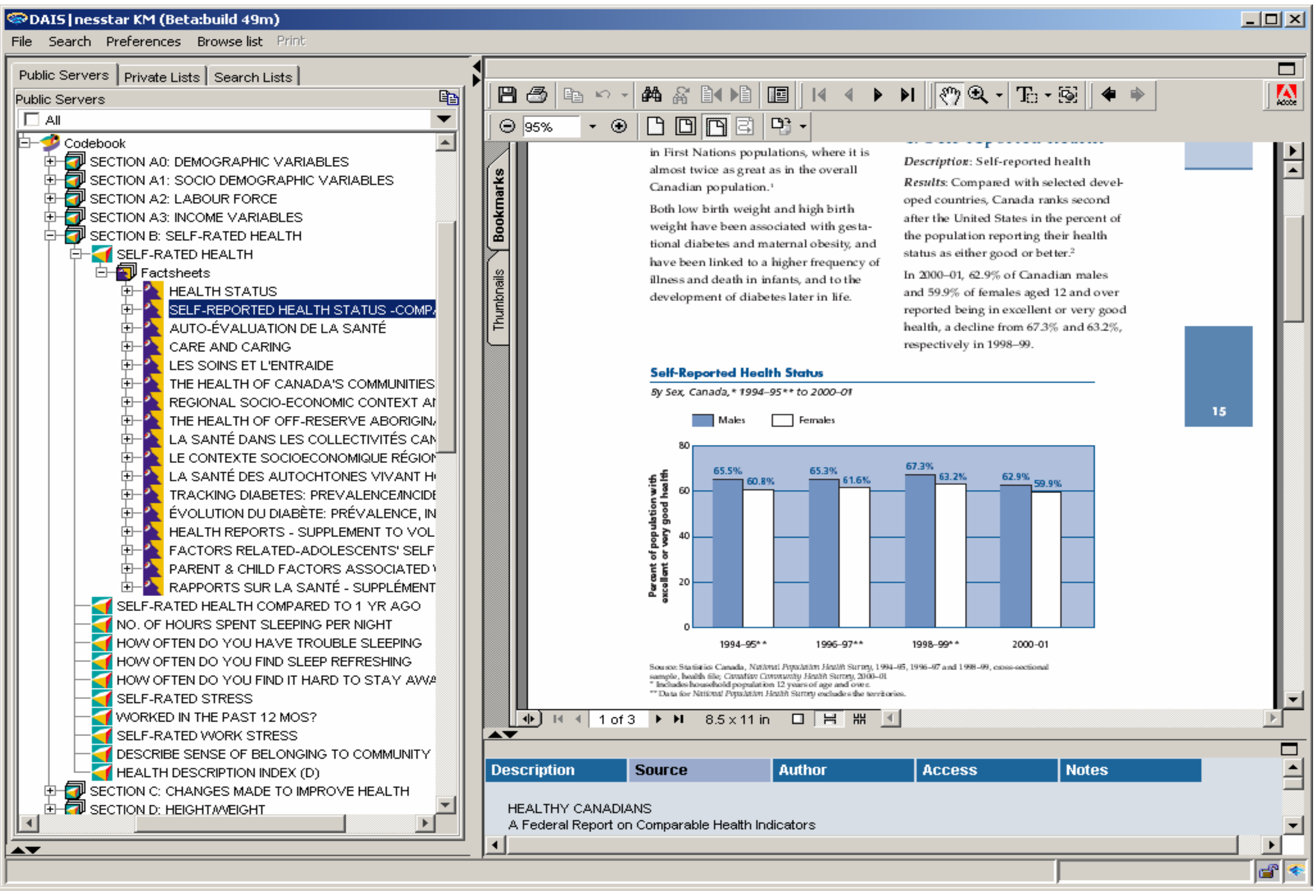
33 There are two areas of production where DDI standardization makes significant contributions – in the planning and management of data expenditures and production and delivery of knowledge.

34 **Planning and management** The information planning and management process is critical for helping to assure that the enormous sums allocated to NSOs are used efficiently to assure maximal policy effectiveness. Health Canada's metabase, standardized across most of the surveys of relevance to the Department since the 1970s, is invaluable as a planning tool. For those at senior levels, being able to ascertain quickly at the desktop whether a question has been asked, where, what the results were, and how it has been used, is invaluable. At working levels the more detailed applications have included **1)** ascertaining information gaps **2)** identifying duplication **3)** comparing the results of and making instrumentation decisions concerning different question wordings **4)** helping to assure comparability of measures across collections, and **5)** helping to decide whether, how often and when to ask a question.

35 Of special use has been DDMS' capability for producing inventories of questions that have been asked on various topics or target groups over the 30-year period, including the associated data items, results and research reports. Topical inventories are extraordinarily useful to program managers and subject matter specialists, who are then able to make solid recommendations for further data collection and analytical investments in their areas.

36 Information co-ordination specialists are able routinely to search for, develop and print groups of questions on topics or concepts that are being considered in survey planning processes. Given the limitations on space in surveys, and the competition among topics, the ability to show whether a question has been used or not in research reports (as in Figure 7) can be very useful in assessing value and tradeoffs.

37 **Knowledge Production** A typical knowledge development paradigm is illustrated below. Policy


DAIS | nesstar KM (Beta:build 49m)
File Search Preferences Browse list Print
Public Servers | Private Lists | Search Lists
Public Servers
All
Codebook
SECTION A0: DEMOGRAPHIC VARIABLES
SECTION A1: SOCIO DEMOGRAPHIC VARIABLES
SECTION A2: LABOUR FORCE
SECTION A3: INCOME VARIABLES
SECTION B: SELF-RATED HEALTH
SELF-RATED HEALTH
Factsheets
HEALTH STATUS
SELF-REPORTED HEALTH STATUS -COMP/
AUTO-ÉVALUATION DE LA SANTÉ
CARE AND CARING
LES SOINS ET L'ENTRAIDE
THE HEALTH OF CANADA'S COMMUNITIES
REGIONAL SOCIO-ECONOMIC CONTEXT AI
THE HEALTH OF OFF-RESERVE ABORIGIN/
LA SANTÉ DANS LES COLLECTIVITÉS CAN
LE CONTEXTE SOCIOECONOMIQUE RÉGION
LA SANTÉ DES AUTOCHTONES VIVANT H
TRACKING DIABETES: PREVALENCE/INCIDE
ÉVOLUTION DU DIABÈTE: PRÉVALENCE, IN
HEALTH REPORTS - SUPPLEMENT TO VOL
FACTORS RELATED-ADOLESCENTS' SELF
PARENT & CHILD FACTORS ASSOCIATED
RAPPORTS SUR LA SANTÉ - SUPPLÉMENT
SELF-RATED HEALTH COMPARED TO 1 YR AGO
NO. OF HOURS SPENT SLEEPING PER NIGHT
HOW OFTEN DO YOU HAVE TROUBLE SLEEPING
HOW OFTEN DO YOU FIND SLEEP REFRESHING
HOW OFTEN DO YOU FIND IT HARD TO STAY AWA
SELF-RATED STRESS
WORKED IN THE PAST 12 MOS?
SELF-RATED WORK STRESS
DESCRIBE SENSE OF BELONGING TO COMMUNITY
HEALTH DESCRIPTION INDEX (D)
SECTION C: CHANGES MADE TO IMPROVE HEALTH
SECTION D: HEIGHT/WEIGHT
95%
Bookmarks
Thumbnails
15
1 of 3
8.5 x 11 in
Description
Source
Author
Access
Notes

in First Nations populations, where it is almost twice as great as in the overall Canadian population.[1]

Both low birth weight and high birth weight have been associated with gestational diabetes and maternal obesity, and have been linked to a higher frequency of illness and death in infants, and to the development of diabetes later in life.

Description: Self-reported health

Results: Compared with selected developed countries, Canada ranks second after the United States in the percent of the population reporting their health status as either good or better.[2]

In 2000–01, 62.9% of Canadian males and 59.9% of females aged 12 and over reported being in excellent or very good health, a decline from 67.3% and 63.2%, respectively in 1998–99.

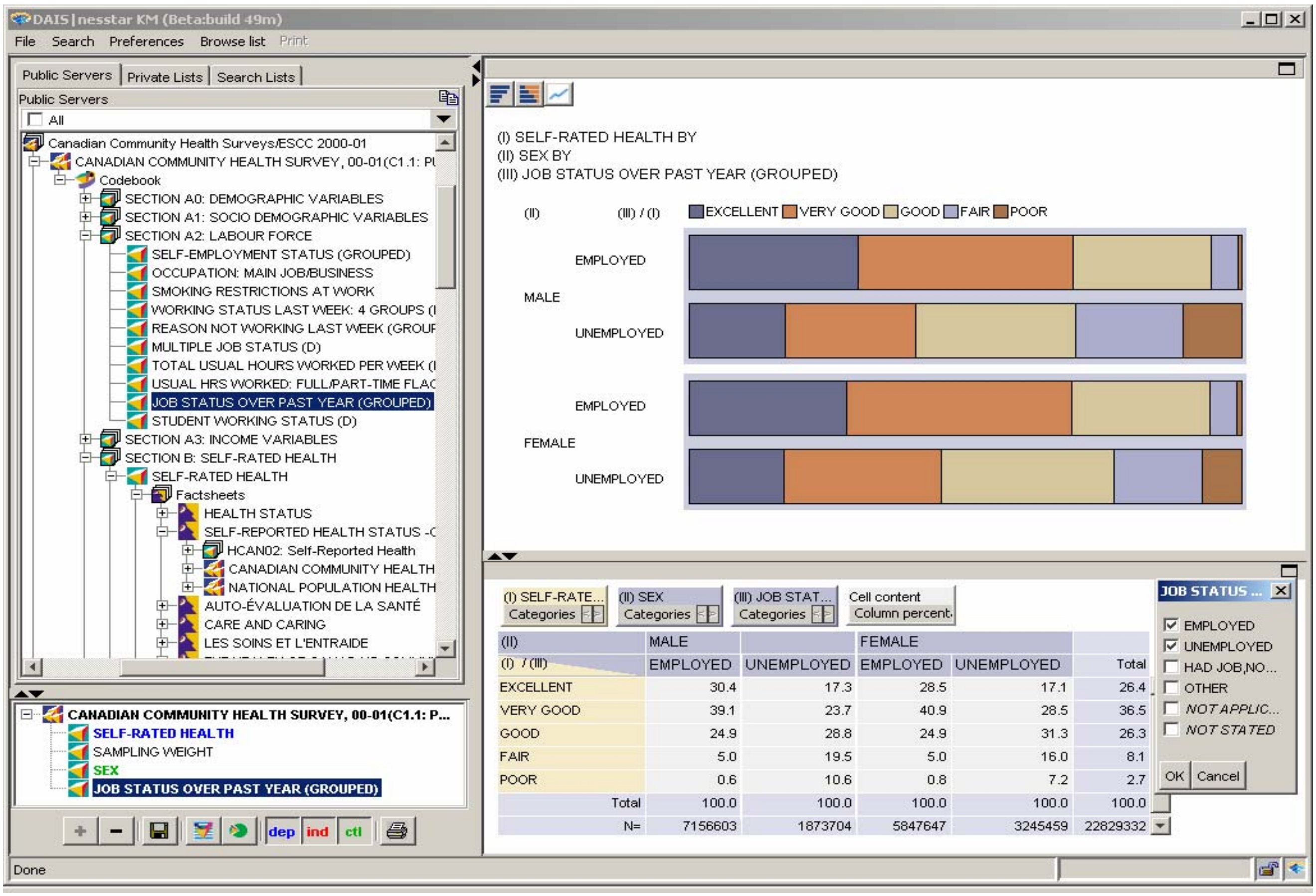
Self-Reported Health Status

By Sex, Canada, 1994–95** to 2000–01*

Source: Statistics Canada, *National Population Health Survey*, 1994–95, 1996–97 and 1998–99, cross-sectional sample, health file; *Canadian Community Health Survey*, 2000–01
* Includes household population 12 years of age and over.
** Data for *National Population Health Survey* excludes the territories.

HEALTHY CANADIANS
A Federal Report on Comparable Health Indicators

Figure 7 ‘Drilling-up’ (linking) from a data item to knowledge products produced from it


DAIS | nesstar KM (Beta:build 49m)
File Search Preferences Browse list Print
Public Servers | Private Lists | Search Lists
Public Servers
All
Canadian Community Health Surveys/ESCC 2000-01
CANADIAN COMMUNITY HEALTH SURVEY, 00-01(C1.1: PU
Codebook
SECTION A0: DEMOGRAPHIC VARIABLES
SECTION A1: SOCIO DEMOGRAPHIC VARIABLES
SECTION A2: LABOUR FORCE
SELF-EMPLOYMENT STATUS (GROUPED)
OCCUPATION: MAIN JOB/BUSINESS
SMOKING RESTRICTIONS AT WORK
WORKING STATUS LAST WEEK: 4 GROUPS (I
REASON NOT WORKING LAST WEEK (GROUF
MULTIPLE JOB STATUS (D)
TOTAL USUAL HOURS WORKED PER WEEK (I
USUAL HRS WORKED: FULL/PART-TIME FLAC
JOB STATUS OVER PAST YEAR (GROUPED)
STUDENT WORKING STATUS (D)
SECTION A3: INCOME VARIABLES
SECTION B: SELF-RATED HEALTH
SELF-RATED HEALTH
Factsheets
HEALTH STATUS
SELF-REPORTED HEALTH STATUS -C
HCAN02: Self-Reported Health
CANADIAN COMMUNITY HEALTH
NATIONAL POPULATION HEALTH
AUTO-ÉVALUATION DE LA SANTÉ
CARE AND CARING
LES SOINS ET L'ENTRAIDE
CANADIAN COMMUNITY HEALTH SURVEY, 00-01(C1.1: P...
SELF-RATED HEALTH
SAMPLING WEIGHT
SEX
JOB STATUS OVER PAST YEAR (GROUPED)
dep ind ctl
Done
(I) SELF-RATED HEALTH BY
(II) SEX BY
(III) JOB STATUS OVER PAST YEAR (GROUPED)
(I) (III) / (I)
EXCELLENT VERY GOOD GOOD FAIR POOR
MALE
EMPLOYED
UNEMPLOYED
FEMALE
EMPLOYED
UNEMPLOYED
(I) SELF-RATE... Categories
(II) SEX Categories
(III) JOB STAT... Categories
Cell content Column percent
JOB STATUS ...
☑ EMPLOYED
☑ UNEMPLOYED
☐ HAD JOB,NO...
☐ OTHER
☐ NOT APPLIC...
☐ NOT STATED
OK Cancel

(II)	MALE		FEMALE		
(I) / (III)	EMPLOYED	UNEMPLOYED	EMPLOYED	UNEMPLOYED	Total
EXCELLENT	30.4	17.3	28.5	17.1	26.4
VERY GOOD	39.1	23.7	40.9	28.5	36.5
GOOD	24.9	28.8	24.9	31.3	26.3
FAIR	5.0	19.5	5.0	16.0	8.1
POOR	0.6	10.6	0.8	7.2	2.7
Total	100.0	100.0	100.0	100.0	100.0
N=	7156603	1873704	5847647	3245459	22829332

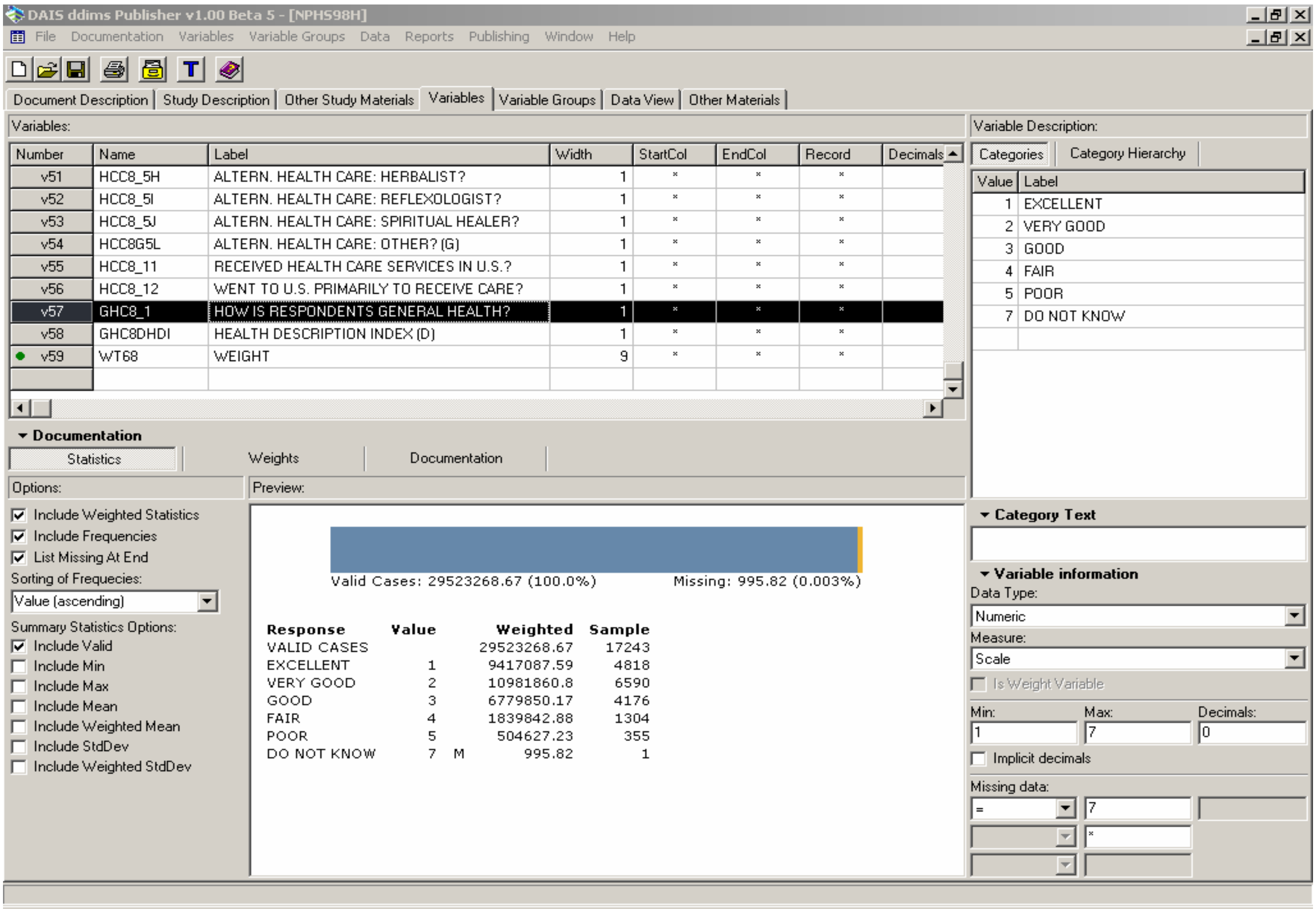
Figure 8 Extending the knowledge – adding labour force status to the data point for 2000 shown in Fig. 7

workers often prefer to start with the existing information or knowledge base, hopefully vetted, which can then be refined with confidence to provide information that relates to their needs. In Figure 6, Section IV above, the user had found and examined a variable of interest, ‘self reported health status’, but then right clicked and chosen “Get Factsheets’ to produce the list of research reports and publications that have employed this variable, shown in Figure 7.

38 In Figure 7 the user is examining reports that show analyses which use the variable of interest. Perhaps there is a meeting with counterparts in the labour component of the Human Resources Department later in the day, and a quick chart demonstrating the relationship between being employed and health status, ideally in an international context, will be helpful. A report is found that presents health status indicators in an international context, but not by employment status. So in Figure 8, below, the analyst has found and dragged the ‘job status’ and gender variables from the codebook to the action window, lower left panel, and requested the analysis shown, in which health status is the dependent variable, job status the independent variable, and gender a control variable. The chart is displayed full screen and cut and pasted for the meeting.

V DDI Metadata in Statistical Production Processes

39 The DDI applies only to the final stages of data production, when data are being prepared for internal analysis or for release to the general public or clients. The DAIS|nesstar DDI Management System (DdiMS), which has been designed following the general functional model in Figure 1, is provided to assist in entering or capturing, editing, managing, printing and exchanging metadata in the DDI XML format. DdiMS is also used to bind data and metadata for publishing in the NSD statistical format used in Nesstar or DAIS|nesstar servers, and for exporting data in the internal formats of popular statistical packages, such as SPSS, SAS and Stata. The main data item entry and editing screen is shown in Figure 9.



Number	Name	Label	Width	StartCol	EndCol	Record	Decimals
v51	HCC8_5H	ALTERN. HEALTH CARE: HERBALIST?	1	*	*	*	
v52	HCC8_5I	ALTERN. HEALTH CARE: REFLEXOLOGIST?	1	*	*	*	
v53	HCC8_5J	ALTERN. HEALTH CARE: SPIRITUAL HEALER?	1	*	*	*	
v54	HCC8G5L	ALTERN. HEALTH CARE: OTHER? (G)	1	*	*	*	
v55	HCC8_11	RECEIVED HEALTH CARE SERVICES IN U.S.?	1	*	*	*	
v56	HCC8_12	WENT TO U.S. PRIMARILY TO RECEIVE CARE?	1	*	*	*	
v57	GHC8_1	HOW IS RESPONDENTS GENERAL HEALTH?	1	*	*	*	
v58	GHC8DHDI	HEALTH DESCRIPTION INDEX (D)	1	*	*	*	
● v59	WT68	WEIGHT	9	*	*	*	

Value	Label
1	EXCELLENT
2	VERY GOOD
3	GOOD
4	FAIR
5	POOR
7	DO NOT KNOW

☑ Include Weighted Statistics
☑ Include Frequencies
☑ List Missing At End
Sorting of Frequecies: Value (ascending)
Summary Statistics Options:
☑ Include Valid
☐ Include Min
☐ Include Max
☐ Include Mean
☐ Include Weighted Mean
☐ Include StdDev
☐ Include Weighted StdDev

```
Valid Cases: 29523268.67 (100.0%)          Missing: 995.82 (0.003%)

Response        Value        Weighted    Sample
VALID CASES               29523268.67     17243
EXCELLENT           1      9417087.59      4818
VERY GOOD           2      10981860.8      6590
GOOD                3      6779850.17      4176
FAIR                4      1839842.88      1304
POOR                5       504627.23       355
DO NOT KNOW         7   M      995.82         1
```

Data Type: Numeric
Measure: Scale
☐ Is Weight Variable
Min: 1 Max: 7 Decimals: 0
☐ Implicit decimals
Missing data: = 7
*

Figure 9 Main variable entry and editing screen in DdiMS showing statistics selection

40 The DDI has not been concerned with the many other uses for metadata in the statistical data production processes. If your agency already has effective standardization disciplines in place across all its programs, it should be a simple matter to map the metadata from internal management systems to the DDI, and then prepare software to generate the XML instances. As discussed below, if your agency has not implemented such disciplines, the author and others (Bradley et al, 2003) recommend that you begin immediately to adapt ISO 11179 (http://metadata-stds.org) and its Corporate Metadata Repository (CMR) (Gillman, 2001) statistical extensions for use in your organisation.

V But we are not permitted to provide access to micro data – only macro data

41 All the more reason to develop and provide access to standardized metadata, where possible, for the microdata upon which the aggregates are based! As discussed in paragraphs 18 and 19, in order to achieve better payoff from the huge investments they make in data, NSOs need to take steps to maximize and deliver the resulting knowledge. Pre-formed aggregates are helpful, but can never provide the customized evidence that is needed to support the countless decisions that are made every day, throughout all the domains of intervention to which the investment should contribute. The fundamental requirement for knowledge producers, and others who advise on policy decisions, is to be able to interact directly with micro data in a hands-on manner, to perform custom analyses within tight time frames. *Alternatively, when the requisite micro data are inaccessible, analysts need to have access to excellent metadata so that they can specify analyses for production on a fast and responsive basis by those who do have the access.*

42 Moreover, a great deal of the process of having custom analyses produced behind the firewall can be automated. The generic data dictionary inherent in the DDI, for example, means that systems like DAIS|nesstar can generate the code for analyses by SPSS, SAS, Stata or other packages. The code can then be submitted automatically for running behind the firewall with the intervention of NSO staff. In the first instance, however, the micro data simply need to be made known through metadata so that users can understand what is there and formulate requests based on their knowledge requirements (Nordbotten, 1993). No NSO can ever anticipate, understand and generate all the analyses that are possible, and that will be needed in the future.

VI Metadata Models and Terminology

43 The DDI's semantic structure and content are described in a tag library manual that is available for downloading from the DDI website at http://www.icpsr.umich.edu/DDI/. Given the importance of ISO/IEC 11179 as the only formally balloted international standard for metadata, the work of Gillman (2001) and colleagues at the U.S. Bureau of the Census in developing the CMR extension for statistical metadata, the take up of 11179 and CMR by other agencies such as Statistics Canada, and the importance to users of having DDI instances produced with data at source, a subcommittee of the DDI, including members of the DAIS|nesstar project, carried out extensive semantic mapping of the DDI, ISO 11179, CMR, and the DDMS/DAIS metadata model. The detailed results of the exercise, which was completed by DAIS|nesstar project staff and vetted by Gillman and Green of the DDI Committee, are available from ICPSR or the author. As with many of the papers cited here, they are also available via DAIS|nesstar at http://DAISnestar.ca/daislight/index.jsp under Factsheets/About DAIS/Background papers.

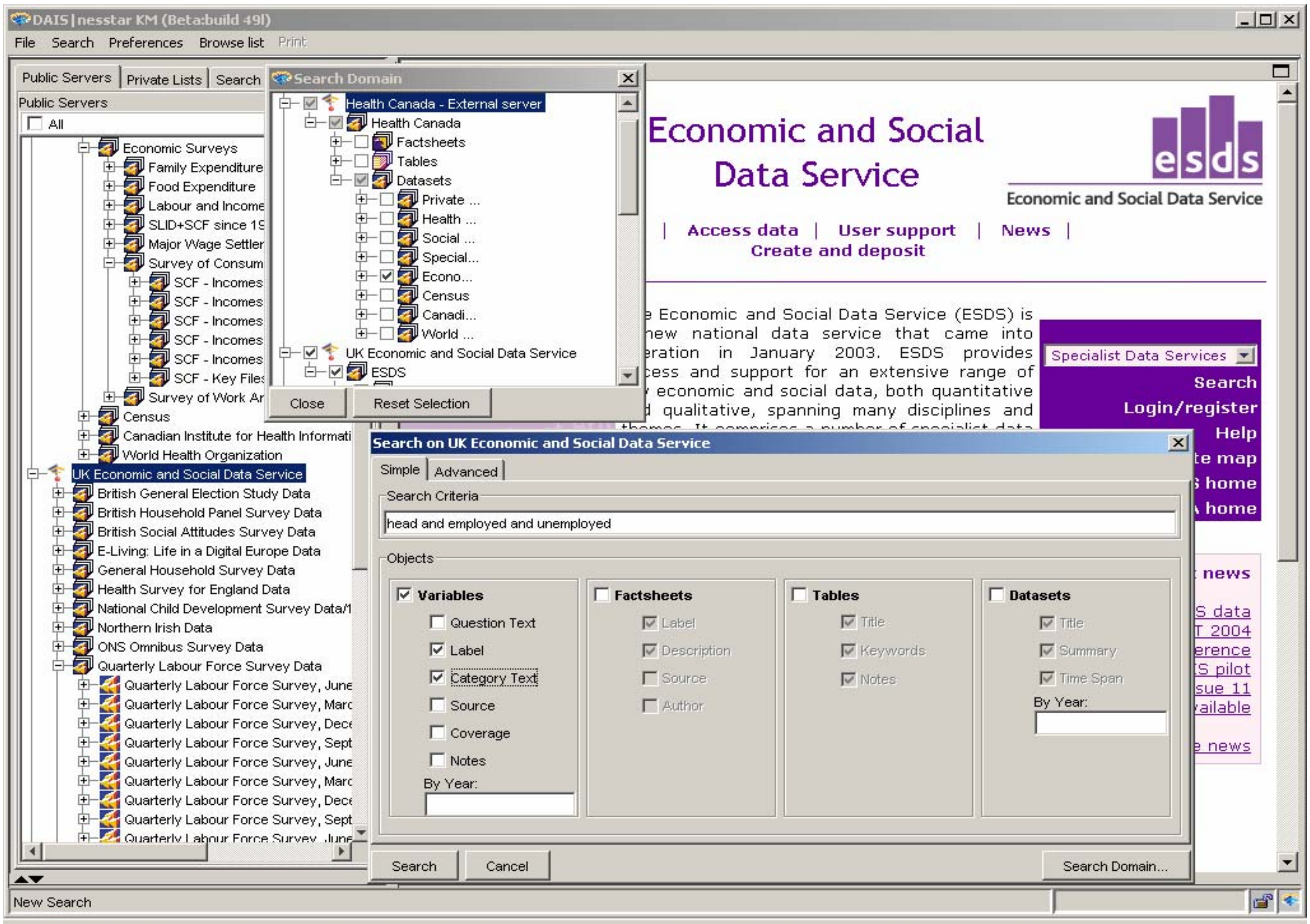
44 The mappings showed an excellent fit between the DDI and CMR. A NSO using CMR should have no problem generating DDI instances from internal systems, provided care is taken to ensure that the adaptation of CMR used includes the DDI fields.

45 The mappings between DDI and ISO 11179 were less congruent, reflecting fundamental differences in objective and approach in the development of the respective standards (Ryssevik et al, 2003). With the collaboration of Gillman, the DAIS|nesstar project has developed an approach to harmonization and a proposed extension to DDI that would enable key elements of 11179 to be captured along with the CMR elements. The extension would strengthen the administration aspects of the DDI as

well as its ability to represent concepts (Ryssevik et al, 2003), (Bradley, Colquhoun and Ryssevik, 2003), two areas in which ISO 11179 is strong and the DDI is weak. The harmonization would be achieved via extensions to a data model representation for the DDI, such as one developed for the DAIS|nesstar project by Hassan and Glover (2001). The development of a formal data model for the DDI, and harmonization with ISO 11179, are now part of the work plan for the DDI Alliance.

VII Searching For, Finding and Using Statistical data in Websites

46 DDI standardisation, coupled with the distributed web server technology devised by Nesstar and extended in the DAIS|nesstar project, enables integrated searches for data items, data sets, tables, fact sheets and research reports. We illustrate with a search at the data item level below. Following the interest in the relationship between health and employment status above, let's suppose that the analyst now wants to extend the indicators in the report in Figure 7 to show trends in employment status and health status in the UK and Canadian populations over time. Figures 10 and 11 show a search for employment status variables in the Canadian Survey of Consumer Finances, a regular annual supplement to the Canadian Labour Force Survey that took place from 1972 to 1997, and in the UK Quarterly Labour Force Survey data sets that are available


DAIS|nesstar KM (Beta:build 49l)
File
Search
Preferences
Browse list
Print
Public Servers
Private Lists
Search
Public Servers
All
Economic Surveys
Family Expenditure
Food Expenditure
Labour and Income
SLID+SCF since 19
Major Wage Settler
Survey of Consum
SCF - Incomes
SCF - Incomes
SCF - Incomes
SCF - Incomes
SCF - Incomes
SCF - Key Files
Survey of Work Ar
Census
Canadian Institute for Health Informati
World Health Organization
UK Economic and Social Data Service
British General Election Study Data
British Household Panel Survey Data
British Social Attitudes Survey Data
E-Living: Life in a Digital Europe Data
General Household Survey Data
Health Survey for England Data
National Child Development Survey Data/1
Northern Irish Data
ONS Omnibus Survey Data
Quarterly Labour Force Survey Data
Quarterly Labour Force Survey, June
Quarterly Labour Force Survey, Marc
Quarterly Labour Force Survey, Dece
Quarterly Labour Force Survey, Sept
Quarterly Labour Force Survey, June
Quarterly Labour Force Survey, Marc
Quarterly Labour Force Survey, Dece
Quarterly Labour Force Survey, Sept
Quarterly Labour Force Survey, June
Search Domain
Health Canada - External server
Health Canada
Factsheets
Tables
Datasets
Private ...
Health ...
Social ...
Special...
Econo...
Census
Canadi...
World ...
UK Economic and Social Data Service
ESDS
Close
Reset Selection
Economic and Social Data Service
esds
Economic and Social Data Service
| Access data | User support | News |
Create and deposit
Specialist Data Services
Search
Login/register
Help
te map
home
home
news
S data
T 2004
erence
S pilot
sue 11
ailable
e news
Search
Cancel
Search Domain...
New Search

e Economic and Social Data Service (ESDS) is new national data service that came into eration in January 2003. ESDS provides cess and support for an extensive range of economic and social data, both quantitative d qualitative, spanning many disciplines and

Search on UK Economic and Social Data Service

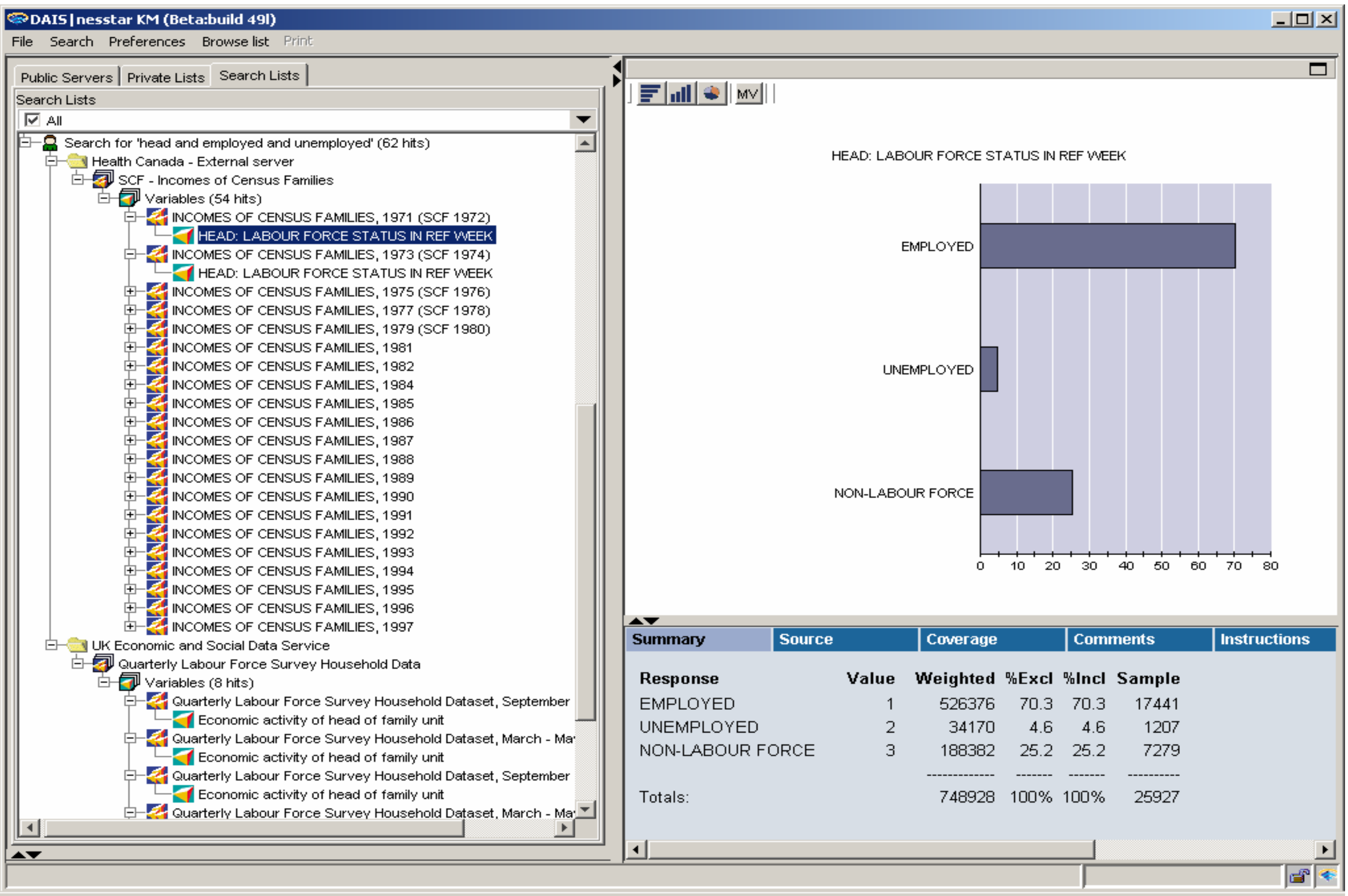
Simple | Advanced

Search Criteria: head and employed and unemployed

Objects

- ☑ Variables: □ Question Text, ☑ Label, ☑ Category Text, □ Source, □ Coverage, □ Notes, By Year:
- □ Factsheets: ☑ Label, ☑ Description, □ Source, □ Author
- □ Tables: ☑ Title, ☑ Keywords, ☑ Notes
- □ Datasets: ☑ Title, ☑ Summary, ☑ Time Span, By Year:

Figure 10 Searching for variables on head's employment status on Canadian and UK servers



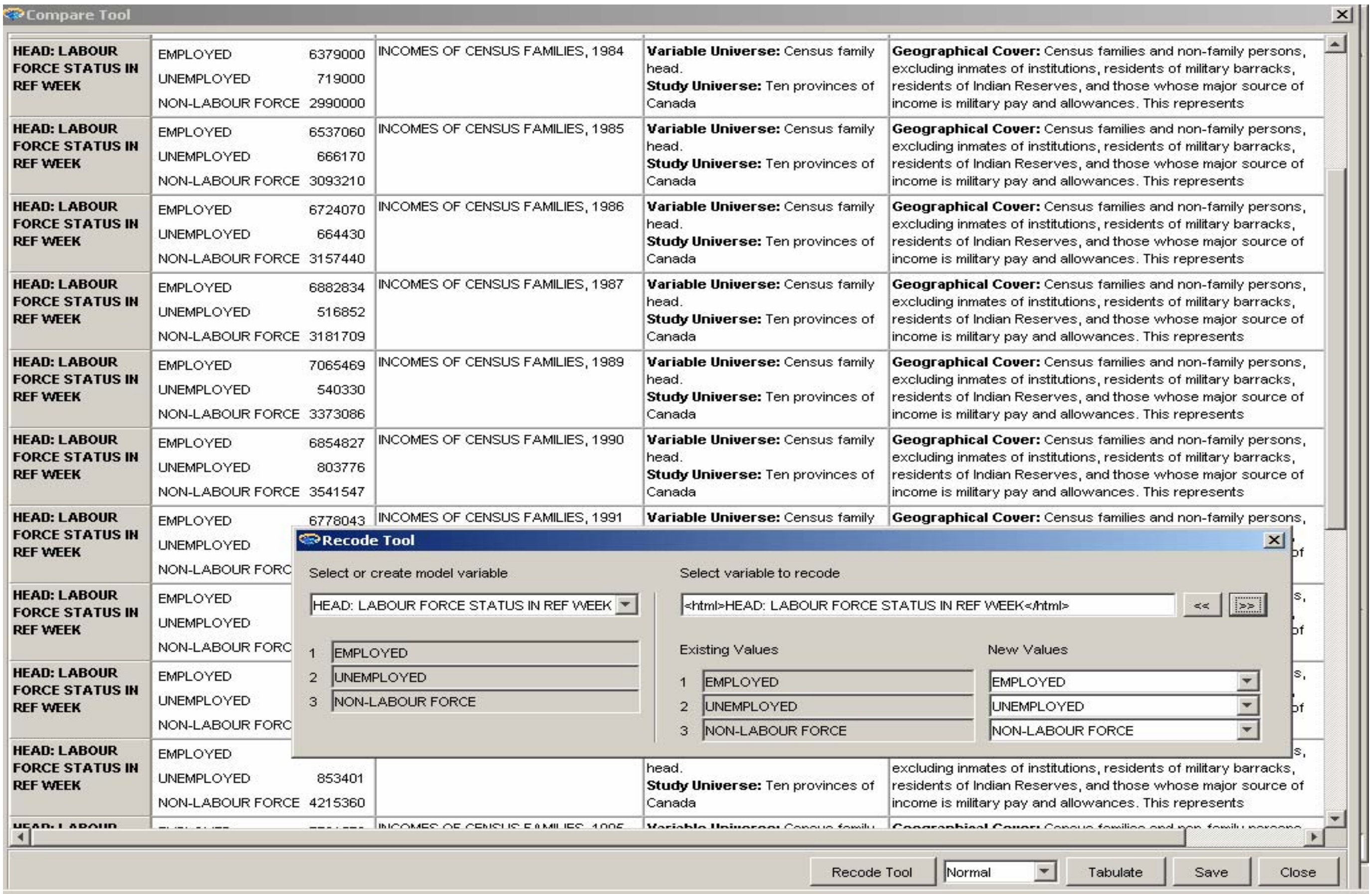
Search Lists

☑ All

- Search for 'head and employed and unemployed' (62 hits)
  - Health Canada - External server
    - SCF - Incomes of Census Families
      - Variables (54 hits)
        - INCOMES OF CENSUS FAMILIES, 1971 (SCF 1972)
          - HEAD: LABOUR FORCE STATUS IN REF WEEK
        - INCOMES OF CENSUS FAMILIES, 1973 (SCF 1974)
          - HEAD: LABOUR FORCE STATUS IN REF WEEK
        - INCOMES OF CENSUS FAMILIES, 1975 (SCF 1976)
        - INCOMES OF CENSUS FAMILIES, 1977 (SCF 1978)
        - INCOMES OF CENSUS FAMILIES, 1979 (SCF 1980)
        - INCOMES OF CENSUS FAMILIES, 1981
        - INCOMES OF CENSUS FAMILIES, 1982
        - INCOMES OF CENSUS FAMILIES, 1984
        - INCOMES OF CENSUS FAMILIES, 1985
        - INCOMES OF CENSUS FAMILIES, 1986
        - INCOMES OF CENSUS FAMILIES, 1987
        - INCOMES OF CENSUS FAMILIES, 1988
        - INCOMES OF CENSUS FAMILIES, 1989
        - INCOMES OF CENSUS FAMILIES, 1990
        - INCOMES OF CENSUS FAMILIES, 1991
        - INCOMES OF CENSUS FAMILIES, 1992
        - INCOMES OF CENSUS FAMILIES, 1993
        - INCOMES OF CENSUS FAMILIES, 1994
        - INCOMES OF CENSUS FAMILIES, 1995
        - INCOMES OF CENSUS FAMILIES, 1996
        - INCOMES OF CENSUS FAMILIES, 1997
  - UK Economic and Social Data Service
    - Quarterly Labour Force Survey Household Data
      - Variables (8 hits)
        - Quarterly Labour Force Survey Household Dataset, September
          - Economic activity of head of family unit
        - Quarterly Labour Force Survey Household Dataset, March - Ma
          - Economic activity of head of family unit
        - Quarterly Labour Force Survey Household Dataset, September
          - Economic activity of head of family unit
        - Quarterly Labour Force Survey Household Dataset, March - Ma

Response	Value	Weighted	%Excl	%Incl	Sample
EMPLOYED	1	526376	70.3	70.3	17441
UNEMPLOYED	2	34170	4.6	4.6	1207
NON-LABOUR FORCE	3	188382	25.2	25.2	7279
		-------------	-------	-------	----------
Totals:		748928	100%	100%	25927

Figure 11 Examining the hits. Search was restricted to Canadian SCF and UK Labour Force Survey



HEAD: LABOUR FORCE STATUS IN REF WEEK	EMPLOYED 6379000 UNEMPLOYED 719000 NON-LABOUR FORCE 2990000	INCOMES OF CENSUS FAMILIES, 1984	**Variable Universe:** Census family head. **Study Universe:** Ten provinces of Canada	**Geographical Cover:** Census families and non-family persons, excluding inmates of institutions, residents of military barracks, residents of Indian Reserves, and those whose major source of income is military pay and allowances. This represents
HEAD: LABOUR FORCE STATUS IN REF WEEK	EMPLOYED 6537060 UNEMPLOYED 666170 NON-LABOUR FORCE 3093210	INCOMES OF CENSUS FAMILIES, 1985	**Variable Universe:** Census family head. **Study Universe:** Ten provinces of Canada	**Geographical Cover:** Census families and non-family persons, excluding inmates of institutions, residents of military barracks, residents of Indian Reserves, and those whose major source of income is military pay and allowances. This represents
HEAD: LABOUR FORCE STATUS IN REF WEEK	EMPLOYED 6724070 UNEMPLOYED 664430 NON-LABOUR FORCE 3157440	INCOMES OF CENSUS FAMILIES, 1986	**Variable Universe:** Census family head. **Study Universe:** Ten provinces of Canada	**Geographical Cover:** Census families and non-family persons, excluding inmates of institutions, residents of military barracks, residents of Indian Reserves, and those whose major source of income is military pay and allowances. This represents
HEAD: LABOUR FORCE STATUS IN REF WEEK	EMPLOYED 6882834 UNEMPLOYED 516852 NON-LABOUR FORCE 3181709	INCOMES OF CENSUS FAMILIES, 1987	**Variable Universe:** Census family head. **Study Universe:** Ten provinces of Canada	**Geographical Cover:** Census families and non-family persons, excluding inmates of institutions, residents of military barracks, residents of Indian Reserves, and those whose major source of income is military pay and allowances. This represents
HEAD: LABOUR FORCE STATUS IN REF WEEK	EMPLOYED 7065469 UNEMPLOYED 540330 NON-LABOUR FORCE 3373086	INCOMES OF CENSUS FAMILIES, 1989	**Variable Universe:** Census family head. **Study Universe:** Ten provinces of Canada	**Geographical Cover:** Census families and non-family persons, excluding inmates of institutions, residents of military barracks, residents of Indian Reserves, and those whose major source of income is military pay and allowances. This represents
HEAD: LABOUR FORCE STATUS IN REF WEEK	EMPLOYED 6854827 UNEMPLOYED 803776 NON-LABOUR FORCE 3541547	INCOMES OF CENSUS FAMILIES, 1990	**Variable Universe:** Census family head. **Study Universe:** Ten provinces of Canada	**Geographical Cover:** Census families and non-family persons, excluding inmates of institutions, residents of military barracks, residents of Indian Reserves, and those whose major source of income is military pay and allowances. This represents
HEAD: LABOUR FORCE STATUS IN REF WEEK	EMPLOYED 6778043 UNEMPLOYED NON-LABOUR FORC	INCOMES OF CENSUS FAMILIES, 1991	**Variable Universe:** Census family	**Geographical Cover:** Census families and non-family persons, of
HEAD: LABOUR FORCE STATUS IN REF WEEK	EMPLOYED UNEMPLOYED NON-LABOUR FORC			s, of
HEAD: LABOUR FORCE STATUS IN REF WEEK	EMPLOYED UNEMPLOYED NON-LABOUR FORC			s, of
HEAD: LABOUR FORCE STATUS IN REF WEEK	EMPLOYED UNEMPLOYED 853401 NON-LABOUR FORCE 4215360		head. **Study Universe:** Ten provinces of Canada	s, excluding inmates of institutions, residents of military barracks, residents of Indian Reserves, and those whose major source of income is military pay and allowances. This represents

Recode Tool

Select or create model variable

HEAD: LABOUR FORCE STATUS IN REF WEEK

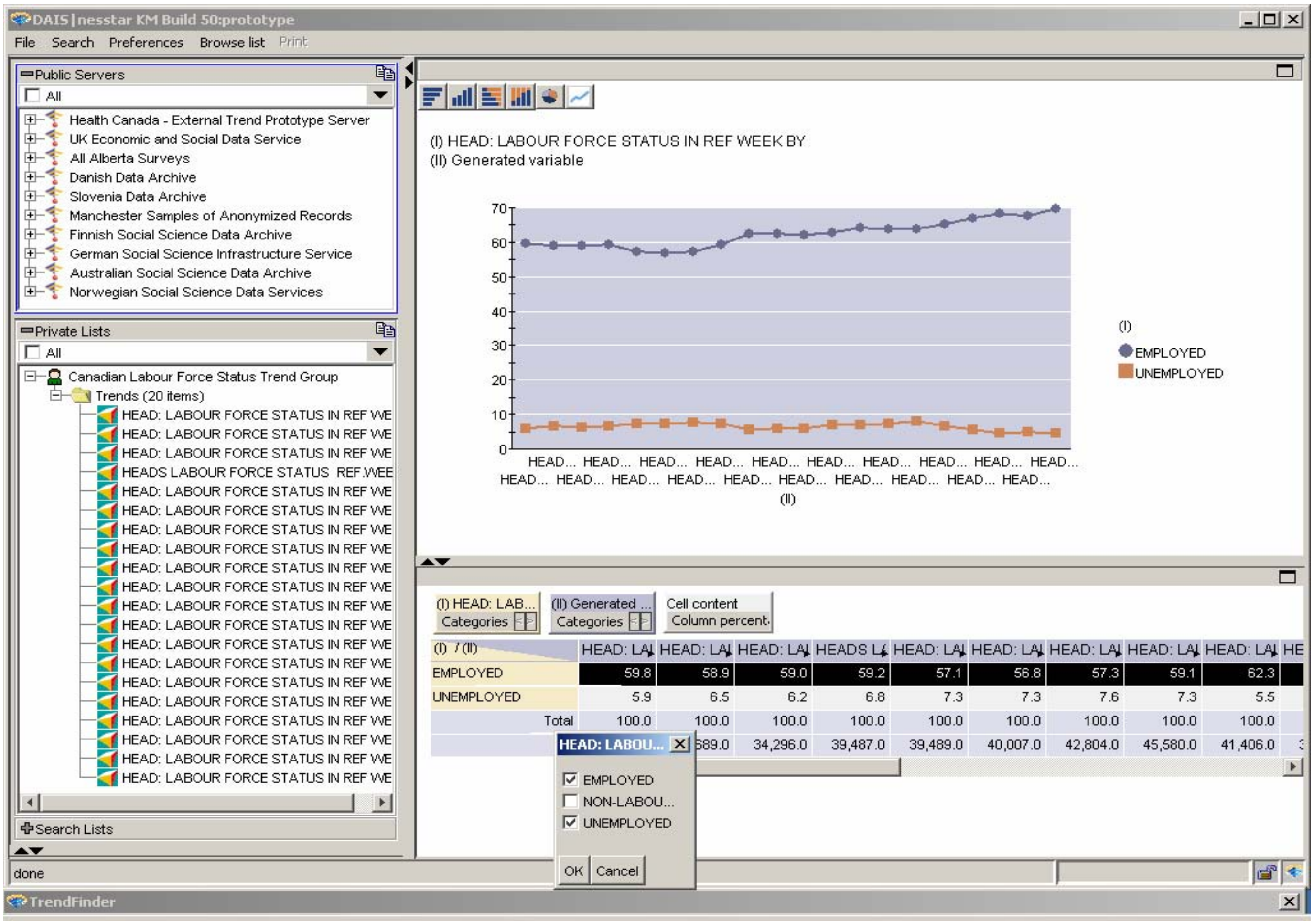
1 EMPLOYED
2 UNEMPLOYED
3 NON-LABOUR FORCE

Select variable to recode

<html>HEAD: LABOUR FORCE STATUS IN REF WEEK</html>

	Existing Values	New Values
1	EMPLOYED	EMPLOYED
2	UNEMPLOYED	UNEMPLOYED
3	NON-LABOUR FORCE	NON-LABOUR FORCE

Figure 12 Creating a trend from the Canadian Survey of Consumer Finance hits


DAIS|nesstar KM Build 50:prototype
File
Search
Preferences
Browse list
Print
Public Servers
All
Health Canada - External Trend Prototype Server
UK Economic and Social Data Service
All Alberta Surveys
Danish Data Archive
Slovenia Data Archive
Manchester Samples of Anonymized Records
Finnish Social Science Data Archive
German Social Science Infrastructure Service
Australian Social Science Data Archive
Norwegian Social Science Data Services
Private Lists
All
Canadian Labour Force Status Trend Group
Trends (20 items)
HEAD: LABOUR FORCE STATUS IN REF WE
HEADS LABOUR FORCE STATUS REF WEE
Search Lists
done
(I) HEAD: LABOUR FORCE STATUS IN REF WEEK BY
(II) Generated variable
70
60
50
40
30
20
10
0
HEAD...
(II)
(I)
EMPLOYED
UNEMPLOYED
TrendFinder

(I) / (II)	HEAD: LA	HEAD: LA	HEAD: LA	HEADS L	HEAD: LA	HEAD: LA	HEAD: LA	HEAD: LA	HEAD: LA	HE
EMPLOYED	59.8	58.9	59.0	59.2	57.1	56.8	57.3	59.1	62.3	
UNEMPLOYED	5.9	6.5	6.2	6.8	7.3	7.3	7.6	7.3	5.5	
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	
		589.0	34,296.0	39,487.0	39,489.0	40,007.0	42,804.0	45,580.0	41,406.0	3

(I) HEAD: LAB...
Categories
(II) Generated ...
Categories
Cell content
Column percent.
HEAD: LABOU...
☑ EMPLOYED
□ NON-LABOU...
☑ UNEMPLOYED
OK
Cancel

Figure 13 Displaying the trend

on the UK Economic and Social Data Service server. Figure 10 shows a portion of the search specification, while in Figure 11 the results of the search are being examined.

47 In Figure 12 the search results from the Canadian server have been placed in the comparison and trend tool. This new facility enables variables from different surveys to be examined on selected metadata fields to determine if measures should be comparable. It also allows code structures and population coverage to be harmonized where possible.

48 The comparable variables are then tabulated across surveys and a cube created to enable the display of trends, or of comparative analyses across subpopulations or geographies. Figure 13 shows the resulting trend, in this case almost a time series, from each of the annual SCF datasets. This development work is being carried out by the Norwegian team, which is also working on the ability to employ comparable control variables in the tabulation for each survey.

49 Hicks (1995) stressed the need to show 'very long trends' across many diverse data sources to support social policy development. We believe that this new facility will enable the display and discovery of trends in multi-dimensional cubes across surveys as never before possible. For the time being, however, the tools are under development, and the numbers generated for Figure 13 are not quite correct – the display is provided for illustrative purposes only and should not be reproduced or cited substantively.

VIII Conclusions

50 In his opening lecture to a Eurostat statistical meta information system workshop held over a decade ago, Nordbotten (1993) described work on preserving metadata, which was carried out as early as the 1960s for the Conference of European Statisticians. 'My experience from national and international statistical

agencies through 40 years indicate that they still don't take enough care of their statistical data treasures in a manner which will satisfy future needs, that they don't agree on and submit to common conceptual and methodological standards which would permit better statistical compatibility, and that they don't take their responsibilities for marketing and disseminating their services and products properly. Users have little knowledge about the content of the statistical data archives, how to combine statistics from different sources and how they could benefit from the large sources of potential information hidden in the data archives. To the extent that producers themselves know the content of their own statistical data sources, the necessary keys to open up the treasures properly for the users are not implemented. These keys are statistical meta-data systems.'

51 Following the difficult times of the 1990s, when many western governments had to implement large cutbacks, funding flows are now being turned back on in priority areas of public concern, such as health. This time, however, there is a sense as never before of the difficult challenges to be faced, and of the limitations in the resources that are available to meet them. Governments increasingly realize that alternatives must be examined rigorously on the basis of available evidence, and that there must be accountability and transparency for the interventions that are made. There is also a growing sense of the global interdependence of risks and interventions, and especially their economic impacts. Users in government departments cannot continue to wait for the deficiencies described by Nordbotten to be corrected.

52 Fortunately, from a technical perspective at least, users should not have to wait much longer. Much progress has been made in metadata standards and systems since 1993. The present paper has described a simple international standard for end user data products, and shown how it can transform applied research and evidence-based policy and program management applications in government. It provides a glimpse of the payoffs for NSOs and their clients that immediate use of a standard like the DDI can engender.

53 Neither the DDI standard nor the tools discussed here are perfect. Much work remains to be done, especially in the areas of harmonization and integration with the other complementary initiatives that are being discussed at the present work session, such as ISO 11179 at the microdata level, and SDMX for aggregate and time-series data. But, we submit, there is enough in the DDI and its tools already that is good enough for NSOs to begin to make investments in describing their microdata using this standard. A general agreement and mission of this nature, and the payoffs that will quickly ensue, might also serve as a means for galvanizing metadata standardization efforts in the other domains of statistical production as well. And surely most data producers would want to publish their own official DDI descriptors, rather than leave it to local support personnel.

54 The meeting may wish to discuss how NSOs can be encouraged to adopt and implement a common metadata standard for end user microdata, and approaches for better harmonization of end-user requirements for metadata with those of internal statistical production systems, ISO 11179/CMR and SDMX.

IX References

Alexander, Cynthia J., **Towards the Information Edge and Beyond: Enhancing the Value of Information in Public Agencies, Report Submitted to Justice Canada**, Department of Political Science, University of Western Ontario, London, Ontario, 1990

Bargmeyer, B., Remarks in **Panel Discussion: Creating a Data Registry – Methods and Tools**, **Metadata Standards for End-User Access to Data**, INFocus 2000 International Health Information Conference, Vancouver, June 2000

Barge, Sylvia J. and Marks, G.A., **OSIRIS III Manual Volume I: System and Program Description**, Release 2 Edition, Ann Arbor Michigan, Institute for Social Research, University of Michigan, 1973

Bradley, W. J., Diguer, John, and Ellis, R. J. E., **Methods for Producing Interchangeable Data Dictionaries and Documentation**, Paper prepared for meetings of the International Association for Social Science Information Service and Technology (IASSIST), Poughkeepsie, N.Y., 1990

Bradley, W. J., Diguer, J., Ellis, R. K., and Ruus, Laine, **DDMS: A PC-Based Package for Managing Social Science Data Dictionaries and Documentation - Introduction and Basic Functions**, 7th Draft Edition, Social Environment Information, Information Systems Directorate, Health and Welfare Canada, April 1991.

Bradley, W. J., Diguer, J., Ellis, R. K., and Ruus, Laine, **DDMS: A PC-Based Package for Managing Social Science Data Dictionaries and Documentation - Reference Manual**, 12th Draft Edition, Social Environment Information, Policy, Planning and Information Branch, Health and Welfare Canada, April 1992

Bradley, W. J., Diguer, J. and Touckley, Lulu, **DDMSS: Data Dictionary Management System Supplemental - User's Guide**, 1st Draft Edition, Social Environment Information, Policy, Planning and Information Branch, Health and Welfare Canada, August, 1992

Bradley, W.J., **Metadata Matters: Standardizing Metadata for Improved Management and Delivery in National information Systems**, Plenary presentation, Annual meetings of International Association for Social Science Information Service and technology, University of Edinburgh, May 1993

Bradley, W. J., Hum, Janine and Khosla, Prem, **Metadata Matters: Standardizing Metadata for Improved Management and Delivery in National Information Systems - Parts 1 Introduction and Overview**, Bureau of Surveillance and Field Epidemiology, Laboratory Centre for Disease Control, Health Canada, 1994

Bradley, W. J., Hum, Janine and Khosla, Prem, **Metadata Matters: Standardizing Metadata for Improved Management and Delivery in National Information Systems - Parts 2 The DDMS/DAIS Model**, Bureau of Surveillance and Field Epidemiology, Laboratory Centre for Disease Control, Health Canada, 1994

Bradley, W. J., Hum, Janine and Khosla, Prem, **Metadata Matters: Standardizing Metadata for Improved Management and Delivery in National Information Systems - Parts 3 Implementing the Model**, Bureau of Surveillance and Field Epidemiology, Laboratory Centre for Disease Control, Health Canada, 1994

Bradley, W. J. and Silins, J. **Building a Virtual Information Warehouse through Standards, Cooperation and Partnerships**, Proceedings of Statistics Canada Symposium '95: From Data to Information - Methods and Systems, November, 1995

Bradley, W.J., Musgrave, S. and Ryssevik, J. **Leveraging the gains: The WebDAIS – Nesstar Project**, Presentation prepared for meetings of the International Association for Social Science Information Service and Technology (IASSIST), Amsterdam, May 2001

Bradley, W.J., Colquhoun, G. and Ryssevik, J., **Integrating ISO 11179 and the DDI in a Single Application**, Presentation to Open Forum on Metadata Registries, Statistics Section, Santa Fe, January 2003

Bradley, W.J., Gillman, D.W., Johanis, P. and Ryssevik, J. **Is your Agency ISO Compliant? Standardizing Metadata for Improved Knowledge Delivery in National Information Systems**, Bulletin of the International Statistical Institute 54th Session Proceedings, Berlin, August 2003

Gillman, D. (2001) **Corporate Metadata Repository (CMR) Model**, Proceedings of Initial MetaNet Conference, Voorburg, Netherlands, April 2001

Glover, T. and Hassan, S., **Specification of Database Design**, Data Systems and Standards Division, Health Canada, January 2002

Musgrave, S. in private conversations with the author at the Data Archive, University of Essex, September 1993

Nordbotten, S. **Statistical Meta-Knowledge and Data**, Invited opening lecture, Conference Proceedings, Statistical Meta Information Systems Workshop, February 1993, EUROSTAT

Roistacher, R. C., **The Data Interchange File: A First Report**, Document No. 207, Center for Advanced Computation, University of Illinois, Urbana, Illinois 61801, 1976

Roistacher, Richard C., **A Style Manual for Machine-Readable Data Files and Their Documentation**, Draft2, Center for Advanced Computation, University of Illinois, Urbana, IL 61801, February 1978

Ryssevik, J. (1999), **Providing Global Access to Distributed Data through Metadata Standardization - The Parallel Stories of NESSTAR and the DDI**, Working Paper no.10 from the UN/ECE Work Session on Statistical Metadata, Geneva, September 1999

Ryssevik, J., Glover T. and Colquhoun, G., **Relationship to ISO 11179,** Data Systems and Standards Division, Health Canada, 2003

Weibel, S. in Addendum to Final Report “Electronic Preservation of Data Documentation: Complementary SGML and Image Capture” SBR-9617813 **Results of the Evaluation of the Data Documentation Initiative (DDI)**, National Science Foundation, Washington, April 2001

Hicks, P., **The Role of Statistics in Making Social Policy**, Presentation to Symposium '95: From Data to Information - Methods and Systems, Statistics Canada, November 1995