



Leveraging the Gains: The WebDAIS-NESSTAR Project

Bill Bradley, Health Canada Simon Musgrave, UK Data Archive Jostein Ryssevik, Norwegian Social Science Data Services

> Bill_Bradley@hc-sc.gc.ca simon@essex.ac.uk jostein.ryssevik@nsd.uib.no

WebDAIS Project

Initiated in January 2001

 Building on and leveraging the DDI, and the DDMS/DAIS and NESSTAR technologies

Strategic approach:

- Don't create competing servers let's consolidate and strengthen the DDI standard to promote wider buy-in, especially by data producers and the official statistics community
 - The field of dreams all data created to standards, fully integrated and immediately accessible. The 'data web'.
- Capture and build on Europe's \$5M (Can.) investment
- Produce immediate results and payoffs for end users

Outline **SALS** Why a standards-based approach From data graveyards to knowledge greenhouses knowledge is what it's all about Goal of the WebDAIS Project distributed capture and access NESSTAR - Open information space enabled by DDI

- Key thrusts for WebDAIS/NESSTAR
- What WebDAIS will look like
- Everyone can build on and use support the standards and join in!.

Why a standards-based approach Metadata standards

- Make data immediately accessible in many formats and processing packages
- Make data portable and exchangeable
- Enable rapid access for decision support
- Eliminate duplication of effort in documenting data for end user access
- Enable users and support personnel to focus on data needs, substance and knowledge, not technical issues
- Enable comparability, analytical integration across data sets, stovepipes and jurisdictions
- Knowledge creation the black gold of the new information age.
 - The way to sell data archiving preservation

Data Graveyards to Greenhouses

- If you want to sell data (.. archiving, preservation, documentation, sharing, national archives and funding, ...) relate it to knowledge.
- No one cares about data in its own right. Knowledge is what it is all about.
- In fact, it can be argued that even knowledge has little value until it is used to make a decision that enhances economic wealth or personal well being
- The DAIS functional model is based on a model of how knowledge is acquired and used to make 'evidence based decisions'



enable full implementation of the Knowledge Acquisition Pyramid

- Desktop access, with rapid movement in all directions through the pyramid
 - Horizontal integration across resources from many different organizations, jurisdictions, producers
 - Searches for variables across data sets.
 - Group comparable variables, measures, concepts
 - Create time series, indicators, comparative analyses
 - Automated table and chart packages
 - Vertical integration of data, information, knowledge
 - Drill-up from a variable to the analyses, research reports, tables in which it is used
 - Drill-down from research reports and tables to the

s the Knowledge Acquisition

- Well demonstrated by Health Canada's DDMS and DAIS systems
 - DDMS system for entering, editing, importing and exporting metadata in a standardized format
 - Used from the mid '80's to prepare and deliver standardized machine readable and hardcopy codebooks. Users have included the Canadian General Social Survey (GSS), Census, post censal survey programs, Canadian Association of Research Libraries and private sector polling firms.
 - DAIS end user client for providing integrated access to microdata and associated tables and reports
 - A standard tool on all 7,000 desktops in Health Canada

Goal of the WebDAIS Project

To create a distributed version of DDMS/DAIS that will work across many autonomous local servers to better serve Health Canada and its partners at all levels of the health system.

- 'Thousands of decisions are made throughout the health sector every day. Our goal is to ensure that they are made on the basis of the best available evidence.' - Denis Gauthier, ADM Health Canada
 - Relevant data, information and knowledge are created everywhere
 - Challenge is to capture and deliver them, when and where created or needed.

 Need a highly distributed approach, with lot's of local autonomy for dealing with ownership,

DAIS and NESSTAR - Points of *departure*

- DAIS might be described as a closed, well-defined and centralised information space. All the resources are stored on one server and handled by a single publishing authority. This makes it possible to provide a complete and detailed resource catalogue (browse-list) and to keep this map updated whenever new resources are added to the system
- NESSTAR on the other hand is designed and implemented for the Internet. It is designed to support an unlimited and fully distributed information space where no single publishing authority has a complete overview of available resources.

NESSIAK - An open information space based on **DDI**

- Works on basis of the dataset as an object
- Background in the distributed catalogue
- Finds and delivers a dataset with accompanying metadata for immediate browsing and downlaoding
- Facilitates creation of flexible hyperlinked information space

...ways of navigating an information space

- Searching: using terms, keywords and conditions to locate resources, either by searching the content of the resources directly or more indirectly by searching their corresponding metadata (Google style)
- Browsing: finding and navigating resources through catalogues, lists, directories etc. (Yahoo style)
- Linking: jumping from one resource to the next through embedded hyperlinks (Web-style). Although catalogues or directories usually will be driven by hyperlinks, there is a basic difference between the top down view of a browse-list and the network view of a hyperlinked information space.

Private bookmarks in NESSTAR







WebDAIS Project - capture the best of both systems

- NESSTAR brings:
 - Distributed server model built with RDF
 - Fully indexed text searching across multiple servers
 - High speed statistical engine
 - Flexible Java client
 - Powerful bookmark 'language'
- DDMS/DAIS brings:
 - Significant experience in metadata creation ('publishing') at source
 - Know-how in delivering to end-users in public sector environment
 - Integrated tables and knowledge products
 - Cross data set 'codebooks' (groups) for creating time-series, comparative analyses, automated report generation capabilities
 - Relational data base technologies

WebDAIS - What it will look like

- Single window of access for *all* information objects data, information, knowledge products
- Shared across many local servers, organizations, jurisdictions, access and control policies, laws and practices *'information spaces'*

 Integrated across information objects (drillup/drill-down) and across information spaces, (drill sideways) ie. through data, information knowledge and across organizations, jurisdictions, archives, users, producers ...
 Let users choose and work at the level of the

Edit Browse Search User Window Help						
rowse			_	_	_	_ 6
New Favorites Search		MV 🔝				
Dais	Concept	Knowladao/I	aformation		Data	
Factsheets	compati	Knowledge/mormation		Uala		
⊕ Private Sector and Polls						and the second se
anadian institute for Health mormation/clis	Object			21		
National Population Health Survey						
🗐 🗐 Special Topics/Sujets spéciaux		The second second second				
Target Groups/Groupes cibles				all		
🗄 🛐 Diseases/Maladies						
⊕ Health and General Surveys						
National Health Expenditures						
🗄 🗐 Survey Questionnaires						a management of the second
🗄 🕣 About Dais - Workshop Slides and Background Papers					<u></u>	
Tables	Group					
		Fartchast	Table	Dataset	Codebook	Variable
	J					
ables Factsheets Tables						
up - Untitled	1					
						NUM 4:57:









New Favorites Search Image: Standard Search Image: Search Image: Search <	Edit Browse Search User Window Help					
New Favorites Bearch Biller Bulkmarry OF RESULTS Cessell Knowledge/Information Data Consult Cessell Cessell Cessell Cessell Package: SAS Windows Cessell Cessell Cessell Dictionary: Cenadol Colume Cessell Cessell Cessell Uclonary: Cessell Cessell Cessell Cessell Cessell Uclonary: Cessell Cessell Cessell Cess	rowse					
SUMMARY OF RESULTS Cessel Knowledge/Information Data Consolid Cassel Knowledge/Information Data Consolid Cassel Knowledge/Information Data Consolid Cassel Cassel Knowledge/Information Data Consolid Cassel Cassel Cassel Cassel Cassel Consolid Cassel Cassel <t< td=""><td>New Favorites Search</td><td>🔁 🗾 🔝 MV</td><td></td></t<>	New Favorites Search	🔁 🗾 🔝 MV				
Image: Second Periods in Smoking B Image: Second Periods in Smoking B Image: Second Periods in Smoking B Image: Second Periods in Smoking B Image: Periods in Smoking B Image: Second Periods in Smoking B Image: Periods in Smoking B Periods in Smoking B Image: Periods in Smoking B Periods in Smoking B Image: Periods in Smoking B Periods in Smoking B Image: Periods in Smoking B Periods in Smoking B Image: Periods in Smoking B Periods in Smoking B Image: Periods in Smoking B Periods in Smoking B Image: Periods in Smoking B Periods in Smoking B Image: Periods in Smoking B Periods in Smoking B Image: Periods in Smoking B Periods in Smoking B Image: Periods in Smoking B Periods in Smoking B Image: Periods in Smoking B Periods in Smoking B Image: Periods in Smoking B Periods in Smoking B Image: Periods in Smoking B Periods in Smoking B Image: Periods in Smoking B Periods in Smoking B Image: Periods in Smoking B Periods in Smoking B Image: Periods in Smoking B Periods in Smoking B Image: Periods in Smoking B Periods in Smoking B <td< td=""><td></td><td>Concept Knowledge/Information</td><td>Data</td></td<>		Concept Knowledge/Information	Data			
Disconstruction CANSIS Extract Wizerd Package: SAS Windows Dictionary: CANADIAN TOBACCO USE MONITORING'99 (W1 - PERSONS) Case: IF ((PROV = 24) OR (DVSST1 = 1)); Dieles Factsheets Tables Output Location: C:\DaisnewidatalExtractt. Image: Construction of the provide	CTUMS99:Trends in Smoking		📣 🔽			
Package: SAS Windows Dictionary: CANADIAN TOBACCO USE MONITORING'99 (WI - PERSONS) Case: IF ((PROV = 24) OR (DVSST1 = 1)), Output Location: C:Daisnew/data/Extract. Up - Unitiled GE OF RESPONDENT (GRO Finit VPE OF SMOKER 1 (DERIVE_) Print Help Egit < Back Pert Estract	DAIS Extract Wizard					
Dictionary: CANADIAN TOBACCO USE MONITORING'99 (W1 - PERSONS) Case: F((PROV = 24) OR (DVSST1 = 1)); Output Location: C:\DaisnewIdata\Extract\ Up - Untitled GE OF RESPONDENT (GRO VEIGHT VARIABLE: SIX MONT VOC CIOARETTES SMOKED F Extract Help Extt < Back Next> Extract Extract Help Extt < Back Next> Extract	Package:	SAS Windows				
Case: IF ((PROV = 24) OR (DVSST1 = 1)); ables Fatsheets Tables Output Location: C:Daisnew.data/Extract/ IPUP - Untitled NGE OF RESPONDENT (GRO VEIGHT VARIABLE: SIX MONT V9C CIGARETTES SMOKED F EXO F RESPONDENT TYPE OF SMOKER 1 (DERIVE_, TYPE OF SMOKER 1 (DERIVE_, TYPE OF SMOKER 1 (DERIVE_, TYPE OF SMOKER 2 (DERIVE) PROVINCE OF SELECTED RESPONDENT WHENGT VARIABLE: SIX MONTH WEIGHT AGE OF RESPONDENT IFINITE IN THE INFORMATION OF THE INFORMATION O	Dictionary:	CANADIAN TOBACCO USE MONITORING'99 (W1 - PERSONS)				
ables Factsheets Tables up - Untitled Run SAS Windows on extract NGE OF RESPONDENT (GRO WEIGHT VARIABLE: SIX MONT VYG CIGARETTES SMOKED F EX OF RESPONDENT TYPE OF SMOKER 1 (DERIVE	Case:	IF ((PROV = 24) OR (DVSST1 = 1));	rbook Variable			
Iup - Untitled ✓ Run SAS Windows on extract NGE OF RESPONDENT (GRO ✓ Print VVG CIGARETTES SMOKED F ✓ Print SEX OF RESPONDENT YPE OF SMOKER 1 (DERIVE	ables Factsheets Tables Output Location:	C:\Daisnew\data\Extract\				
Neight Variable: SiX MONT Veight Variable: SiX MONT VG CIGARETTES SMOKED F SEX OF RESPONDENT YPE OF SMOKER 1 (DERIVE	up - Untitled	Run SAS Windows on extract				
VG CIGARETTES SMOKED F SEX OF RESPONDENT TYPE OF SMOKER 1 (DERIVE	AGE OF RESPONDENT (GRO WEIGHT VARIABLE: SIX MONT	E Print				
TYPE OF SMOKER 2 (DERIVED) PROVINCE OF SELECTED RESPONDENT WHERE DO YOU USUALLY GET CIGARETTES WEIGHT VARIABLE: SIX MONTH WEIGHT GE OF RESPONDENT I I I I I I I I I I I I I I I I I I I	AVG CIGARETTES SMOKED F BEX OF RESPONDENT TYPE OF SMOKER 1 (DERIVE	<u>H</u> elp E <u>x</u> it < <u>B</u> ack <u>N</u> ext > <u>Extra</u>	act			
	TYPE OF SMOKER T (DERIVED) PROVINCE OF SELECTED RESPONDENT WHERE DO YOU USUALLY GET CIGARETTES WEIGHT VARIABLE: SIX MONTH WEIGHT AGE OF RESPONDENT I I I I I I I I I I I I I I I I I I I					



<u>E</u> dit <u>G</u> roup <u>B</u> rowse <u>S</u> earch <u>U</u> s (Variable) - TYPE OF SMOKER	er <u>W</u> indow <u>H</u> elp 2 (DERIVED)			
Summary Source	Coverage Weight	Comments	Interviewer	
erived variable for type	2 smoking status.		<u> </u>	ED)
lesponse	Value Weight	ed %Excl %Incl	Sample	Current daily
Current smoker Current occasional smoker Cormer daily smoker Cormer occasional smoker Cast experimenter Cifetime abstainer Not stated Total:	1 49653 2 10335 3 55296 4 7518 5 38798 6 79013 99 M 1173 241789	31 20.6 20.5 62 4.3 4.3 02 23.0 22.9 99 3.1 3.1 19 16.1 16.0 69 32.8 32.7 90 0.5 72 100% 100%	2545 670 1846 236 1804 3939 124 11164	30 Current occasional 28 Former daily 26 Former occasional 26 Past experimenter smoker 24 Lifetime abstrainer 22 20 18 16 14 12
Dataset Codebo	ook <u>V</u> ariable Groups	<u>T</u> ables	<u>F</u> actsheets	
<u>H</u> elp <u>G</u> o	to DVSST2	Copy to Group	🕤 (Untitled Fact Grou	up)
ables Factsheets Tables		Summany NOTE: New deriv Smoking status 1 Current daily 2 Current occas 3 Former daily 4 Former occas: 5 Experimental 6 Lifetime abst 9 Not stated History: Survey	SUMMARY OF SOMMAIRE DE TRENDS IN SM PRÉVALENCE SURVEY ESTIN OVERALL CON COMPARABILI FAITS SAILLAN PRÉSENTATIO CONCLUSION KEEPING TRA AU FIL DE L'AC SUMMARY OF RESU	RESULTS ES RÉSULTATS MOKING E DE L'USAGE DE LA CIGARETTE, 1985-1999 MATES NCLUSIONS ITÉ DES ENQUÊTES : MÉTHODOLOGIE ET QUEST NTS ET VUE D'ENSEMBLE ON DES ESTIMATIONS D'ENQUÊTE NS GÉNÉRALES ACK CTUALITÉ
Source: CANADIAN TOBACCO USE MONITORING SURVEY				
			⊻iew	Dataset Variable Groups Tables NUM 6:07:

Some Key Thrusts and Issues

- Metadata publishing the DDMS replacement
- Take up by health partners and data producers
- Integration of DDI with ISO 11179
- Tables DTD and aggregate data methodology.
- The virtual information space common browse lists and views across servers and organizations



Everyone can use

- Health Canada doesn't have a mandate to disseminate and support the technology for other users, but our partners in the project will have the rights to do so.
- Built upon international standards
- The critical issue is creating the metadata the new data publisher will be key
- Designed for widespread distribution and integration
- Support the standards and join in!!