

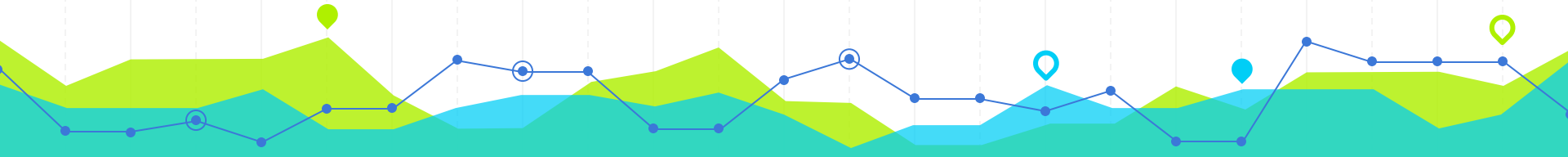


`<odesi>`: supporting different DDI (flavours) from two repository systems

Amber Leahey & Kevin Worthington
April 8th 2016

What is <odesi>?

- Established in 2007, <odesi> is a social science survey data repository that allows users to search, explore, and download data;
- Collaborative service, where DDI-markup and data publishing / maintenance is provided by multiple libraries.





Service provider

- <odesi> is a service of the Ontario Council of University Libraries (OCUL)
- Scholars Portal is the service provider for OCUL <http://scholarsportal.info/>

Books

Journals

GeoPortal

<odesi>

ACE

RACER

Dataverse



Data collections in <odesi>



- Government and commercial data sources (Statistics Canada, other government, IPSOS, Gallup, etc.)
- Aggregate across external providers (ICPSR, CORA, RDC-Master files, and- soon Dataverse)

- High-level of library data curation
- Nesstar Publisher / Web Server driven (DDI 1.2 / Codebook)

ICPSR



GALLUP®




<odesi> search

Basic search

<odesi> Français

Browse  Contact My List 


Find data:


Enter Search Term Anywhere 


Date:


Collections:

☒ All of <odesi>


☒ Statistics Canada (Microdata) 


☒ Statistics Canada (Aggregate data) 


☒ Public Opinion Polls 


☒ Other Data 


Additional Collections:

☐ Statistics Canada (Master File) 

☐ CORA 

☐ ICPSR 

☐ Scholars Portal Dataverse 

 Search

<odesi> contains over 3,535 datasets for the social sciences, with more than 13,000 additional dataset descriptions available for searching.

[Learn more](#)

Learning Tools

[Video Tutorials](#)

[User Guide](#)

[FAQ](#)

About

[About <odesi>](#)

[Who can Access it](#)

[Terms of Use](#)

New Data

ABC News Post-Election Poll #1, November 2000
2010-01-02

ABC News/Washington Post Poll, March 1996
2015-01-02

Multiple Cause of Death Public Use Files, 2000-2002
2010-01-02


Eurobarometer 74.1: Poverty and Social Exclusion, Mobile Phone Use, Economic Crisis, and International Trade,
August-September 2010

Advanced search

Collections

<odesi> Français


Find data:


Enter Search Term 


Date:


Collections:

☒ All of <odesi>


☒ Statistics Canada (Microdata) 


☒ Statistics Canada (Aggregate data) 


☒ Public Opinion Polls 


☒ Other Data 


Additional Collections:

☐ Statistics Canada (Master File) 

☐ CORA 

☐ ICPSR 

☐ Scholars Portal Dataverse 

 Search

Anywhere

Anywhere

Variable/Category Label

Title

Title (Exact)

Keywords

Abstract

Series

Canadian Community Health Survey, 2012: Annual Component

Explore & Download 

[Abstract](#)


[Study Description](#)

[Documentation](#)

[All Survey Variables](#) **1381**

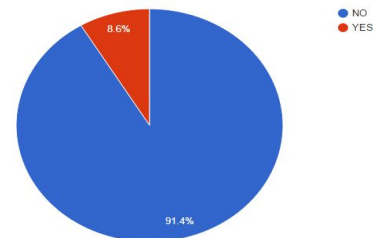
Matching Variables: 66

[+ Add to My List](#)

ID		Label 	Valid Cases	Missing Cases	Minimum	Maximum	Categories
38	CIH_1	Did something to improve health	24909	36798			View
39	CIH_2	Most important change to improve health	14365	47342			View
41	CIH_4	Most important thing to improve health	16761	44946			View
69	CIH_8J	Health improvement - quit smoking	12990	48717			View
77	CCC_031	Has asthma	61630	77			View

Variable: Has asthma





Literal Question

Do you have asthma?

Values	Categories	N ▼	NW
2	NO	56325	27077698.39
1	YES	5305	2384461.58

Summary Statistics

Cases	N	NW
Valid cases	61630	29462159.97
Missing cases	77	28870




Universe

All respondents


Notes

<odesi> repository


<odesi> Français

Browse  Contact  My List 

Canadian Community Health Survey, 2014: Annual Component

Explore & Download 

Abstract Study Description Documentation All Survey Variables **1129** Matching Variables **5** + Add to My List

ID		Label 	Valid Cases
80	CCC_031	Has asthma	63436
81	CCC_035	Asthma - had symptoms/attacks	5379
82	CCC_036	Asthma - took medication	5388
938	WTM_02	Req. vis. med. spec. - type of condition	218
1053	LOPG050	Chronic condition - (G)	4139

Dataset: Canadian Community Health Survey, 2014

Annual Component

Download

Please select a data format from the drop-down box.
If you wish to download a subset of the data, click on the
Click on 'Download' to start downloading.
Please note that you may be asked for a password.

✓ -- Select Data Format --

SPSS
SPSS Portable
Stata v.8
Stata v.7
Nesstar Publisher
NSDstat
Statistica
DIF
DBase
Textfile
Delimited
SAS
Comma Separated Value file


DOWNLOAD SUB

ation

om the drop-down box
ading.

ht analysis results before r
e applying the weights. To
ata collectors and the data

mandé de pondérer les résu


attention particulière, veuillez toujours consulter les procédures de pon
pondérations, sélectionner l'icône de poids  et choisir la variable de
minutieuse. Les personnes charoées de la collecte et de la production des données

Dataset: Canadian Community Health Survey, 2014: Annual Component

Annual Component

Asthma - had symptoms/attacks: Categories 

Asthma - took medication: Categories

Type: Column percentage 

Asthma - had symptoms/attacks	YES	NO	Total
Asthma - took medication			
YES	93.3	64.3	78.4
NO	6.7	35.7	21.6
Total	100.0	100.0	100.0
N=	2,615	2,760	5,375

Copyright (c) Statistics Canada

<odesi> publishing workflow

Data release / added to loading list

Data released from producer (e.g. Statistics Canada)

Data acquired & cleaned / harvested

Data is downloaded, prepared, or harvested from another provider.

Data published

Published to Nesstar / loaded to MarkLogic for searching.

DDI markup

Description of the study, variables, data files, etc.



Flavour of DDI (<odesi>/Nesstar)

- Rich, study level descriptions (library curated)
- Consistent, best practices driven
- Variable-level documentation, question text, universe notes, etc.
- Weighted frequencies
- Dataset / Study 1:1
- Mainly tabular survey data (SPSS, SAS)
- Some aggregate data, described at the study-level only
- HTML tags in XML fields



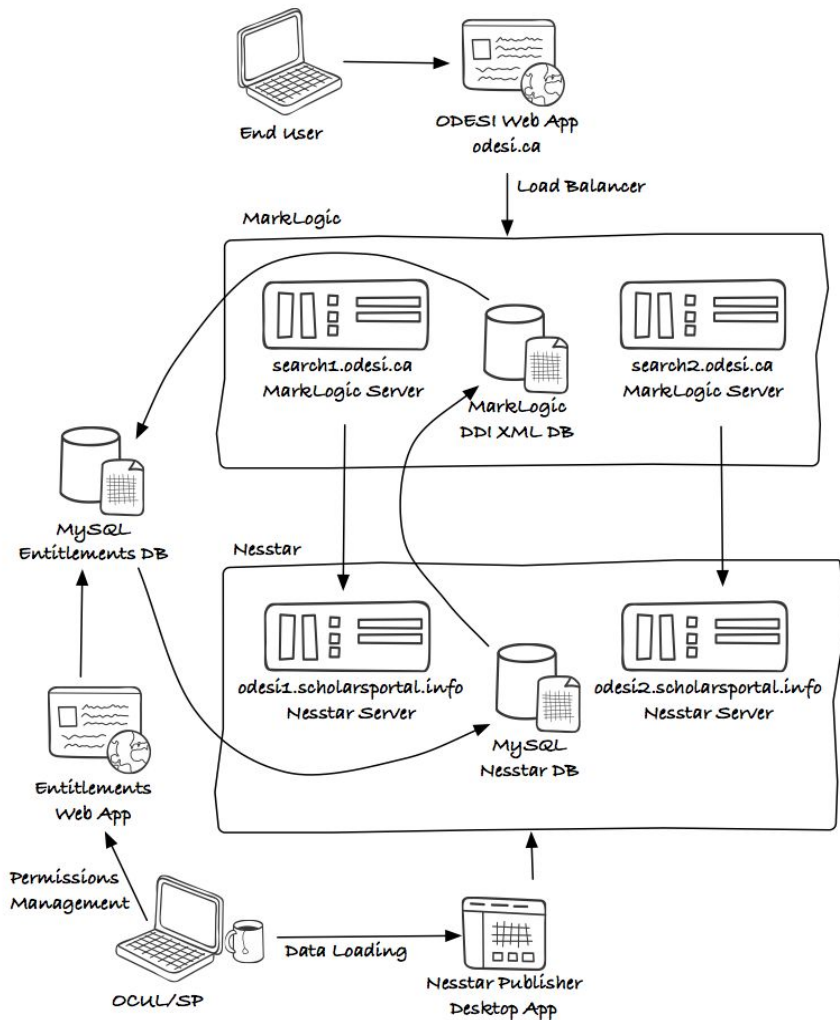
Technical overview

- Frontend

- AngularJS
- Google Charts

- Backend

- MarkLogic w/ REST API (DDI-ish json!)
- MySQL
- NodeJS



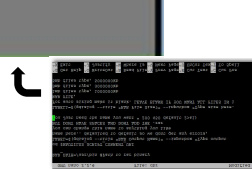
Harvesting / ingesting (Nesstar)

<odesi>

Collections / ML

Nesstar

- <odesi>
- CORA, RDC



Shell script run nightly

1.  Download DDI

2.  Edit

3.  Upload

4.  Update

5.  Discover

Nesstar harvesting / ingesting steps

1. Harvested using Windows directory
2. `codeBook>docDscr>titlStmnt<`
`<year>2008</year>`
`<slang>en</slang>`
3. Ingested with uri prefix
4. Entitlements checked - Append collection
5. Search, Visualize, Access

Scholars Portal Dataverse

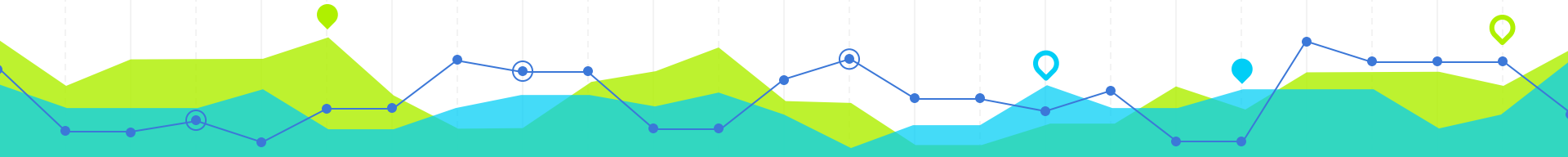
- Harvard IQSS Project
- Since 2012, SP Dataverse offers users online data deposit, management, and sharing platform
- Self-deposit model, anyone in the world can sign up
- 73 Dataverses, 436 Studies, 5 528 Files = approx. 170GB size



Dataverse collections

- 13 OCUL schools (fall 2015)
- Multi-disciplinary
- Citation description (title, author, producer, handle/Doi)
- Support for any file format (Excel, SAS)
- Tabular data ingest, subsetting (TAB file)

AECOM Canada Ltd., "City of Kingston 2008 Travel Survey", <http://hdl.handle.net/10864/11425> UNF: 5:q+38QPEabvSUTlcHrPaVcg== Data Services [Distributor] V1 [Version]



Dataverse search & repository (3.6)



SP Dataverse Network >

Western University Research Data Repository Dataverse

This dataverse houses research data sets created by researchers at Western University (Western Libraries).

Western University

Sort By:

Canadian Fertility Survey, 1984

by Balakrishnan, T.R.; Krótki, Karol J.; Lapierre-Adamcyk, Evelyne

Description: The Canadian Fertility Survey (CFS 1984) was conducted in April-June 1984. The and Humanities Research Council of Canada and executed by co-investigators from three u

Federal Elections in Ontario (1867-1911): Voting and Census Data file

by Kerr, Donald G.; UWU Social Science Computing Laboratory

Description: The Federal Elections in Ontario (1867-1911): Voting and Census Data file was dealing with the effect of religious and ethnic background on voting patterns in Canada. Since



UNIVERSITY OF ALBERTA
LIBRARIES

[data management @ uofa](#)

[about dataverse](#)

[contact us](#)

UAL DATVERSE NETWORK

ALBERTA SURVEY, 1994

doi:10.7939/DVN/10555UNF:5:PKi4F7LpFc2Uz+FdmwCtnQ==

Version: 4 – Released: Mon Jan 18 10:02:33 MST 2016

CATALOGING INFORMATION

[Data & Analysis](#)

[Comments \(0\)](#)

[Versions](#)

POWERED BY THE **Dataverse Network** PROJECT
v. 3.6.2



[Create Account](#)

[Log In](#)

Dataverse search and repository (4.x)

[About](#)[Guides ▾](#)[Support](#)[Sign Up](#)[Log In](#)

Scholars Portal Fun Data Dataverse (SP)

Root Dataverse > Scholars Portal Fun Data Dataverse > **Interesting data**



Interesting data

Leahey, Amber, 2015, "Interesting data", <http://dx.doi.org/10.5072/FK2/ZZVSTW>, Root Dataverse, V2

If you use these data, please add this citation to your scholarly resources. Learn about Data Citation Standards.

Description	test for SAS, FITS
Subject	Mathematical Sciences; Physics; Social Sciences; Other; Astronomy and A
Notes	test

[Files](#) [Metadata](#) [Terms](#) [Versions](#)

Find

5 Files



acisf01873N003_full_img2.fits

FITS - 888.7 KB - Nov 6, 2015 - 0 Downloads

MD5: f9b7b6ca5ac73b0a0e3048b11e0cbf65;

FITS file, 7 HDUs total: The primary HDU; 6 Table HDU(s) 1 Image

HDU(s); The following recognized metadata keys have been found

in the FITS file: INSTRUME, NAXIS0, NAXIS1, TELESCOP, DATE-

OBS, CRVAL2, NAXIS, OBJECT, CRVAL1, EXPTIME;

[À propos](#)[Guides_fr ▾](#)[Soutien](#)[English](#)[French](#)[Spanish](#)

Dataverse Admin ▾

Root Dataverse > **Nouveau dataverse**

Dataverse *

Dataverse

Identifiant *

Catégorie *

Sélectionner un/lune...

Courriel *



Affiliation

Dataverse hôte
/Héberger
Dataverse?

Root

Description

Ce champ prend seulement en charge certains/certaines ??

Champs de métadonnées

Sélectionner les champs de métadonnées à utiliser dans les modèles d'ensemble de données et au moment d'ajouter un ensemble de données à ce dataverse.

- ☒ Utiliser les champs de métadonnées de Root Dataverse
- ☒ Métadonnées de la référence bibliographique (Obligatoire) [\[+\] Voir les champs](#)
- ☐ Métadonnées géospatiales [\[+\] Voir les champs](#)
- ☐ Métadonnées sur les sciences sociales et les sciences humaines [\[+\] Voir les champs](#)
- ☐ Métadonnées liées aux sciences de la vie [\[+\] Voir les champs](#)
- ☐ Métadonnées liées à la revue [\[+\] Voir les champs](#)

Data publishing workflow (Dataverse)

Researcher collects data for study

Collects data in various formats

Researcher creates dataverse/study

Data is described, generate citation

Upload data

Tabular data ingested as tab; DDI file-level generated

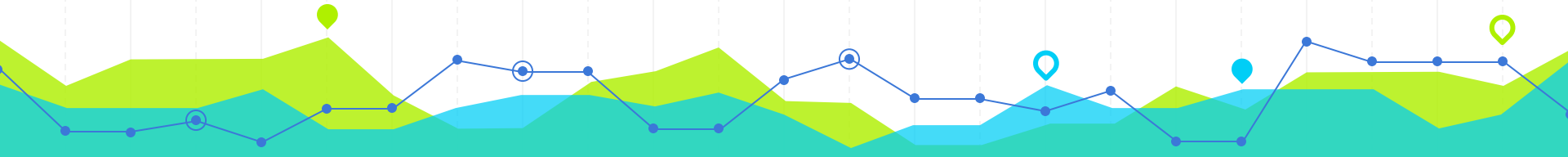
Data published

Published to dataverse network, loaded to MarkLogic for searching



Flavour of DDI (Dataverse)

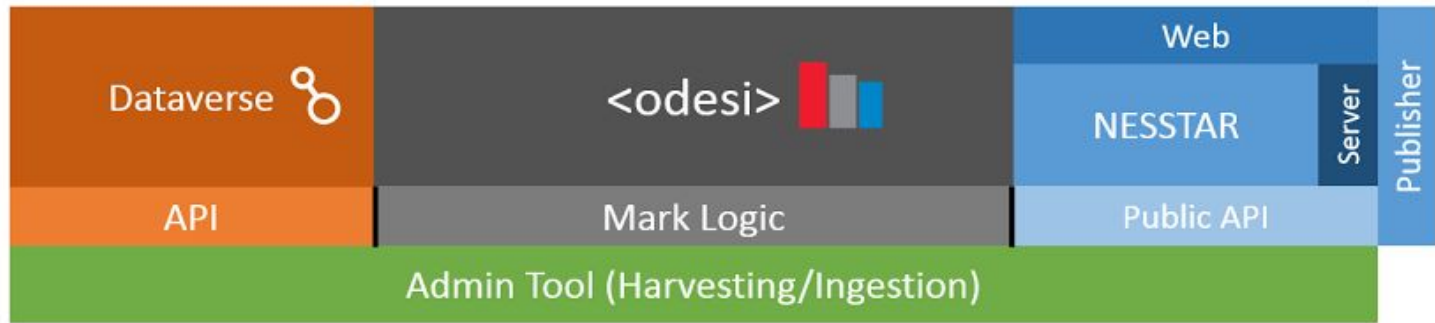
- DDI study-level (Citation, Social Sciences metadata block)
- “Subsettable” files generate the DDI variable-level only
- Support for the description of more formats
- Tab separated files (TAB, .tsv)



DDI comparison: Nesstar & Dataverse

	Nesstar Publisher	Dataverse
DDI Codebook (study-level)	Yes (1.2 compliant)	Yes (DDI-lite)
DDI Codebook (variable-level)	Yes	Some, for tabular files (TAB), no question-level editor (questions, universe, etc.)
Weighting (weighted frequencies)	Yes	No
Datasets : study ratio	Many : many	Many : many
Persistent Identifier	Manual: Any URI (1)	Automated: Handle, DOI, UNF, versioning, other URI (manual)

Harvesting / ingesting (Dataverse)



Dataverse harvesting / ingestion steps

1.  **Download DDI**

2.  **Edit**

3.  **Upload**

4.  **Discover**

1. Harvested using Dataverse API 3.6-4.2
- using `url_search="metadataSearch/dvOwnerId:219"` in SQL DB tbl VDC
2. Inclusion of proprietary tags `syear`, `slang`
3. Ingested with uri prefix
4. Search, Visualize, Access



Study Databases

Select a Database: Dataverse Alberta

URL:

Path:

Tag (XML Node):

Search String (comma delimited):

Last Harvest:

Folder path+done/:

Collection:

URI Prefix:

Add Entitlements: ☐

Last Upload:

Harvest Upload

id	title	status	timestamp
doi-10-7939_DVN_10003	Alberta Survey, 201...	match	2016/01/18...
doi-10-7939_DVN_10004	Alberta Survey, 201...	match	2015/11/04...
doi-10-7939_DVN_10006	Alberta Survey, 201...	match	2015/11/04...
doi-10-7939_DVN_10007	Alberta Survey, 201...	match	2015/11/04...
doi-10-7939_DVN_10008	Alberta Survey, 2010	match	2015/11/04...
doi-10-7939_DVN_10009	Alberta Survey, 2009	match	2015/11/04...
doi-10-7939_DVN_10013	Alberta Survey, 2008	match	2015/11/04...
doi-10-7939_DVN_10546	Alberta Survey, 1999	match	2015/11/04...
doi-10-7939_DVN_10561	Alberta Survey, 1991	match	2016/01/18...

MarkLogic Database

Collection:

Search:

URI	Title	Last Modified Date	Collections
/ALB_DV/doi-10-7939_D...	All Alberta Survey, 1989	04/04/2016 15:1...	new http://scholarsportal.inf...
/ALB_DV/doi-10-7939_D...	All Alberta Survey, 1988	04/04/2016 15:1...	new http://scholarsportal.inf...
/ALB_DV/doi-10-7939_D...	Alberta Survey, 2007	04/04/2016 15:1...	new http://scholarsportal.inf...
/ALB_DV/doi-10-7939_D...	Edmonton Area Survey, 1986	04/04/2016 15:1...	new http://scholarsportal.inf...
/ALB_DV/doi-10-7939_D...	Alberta Survey, 1996B	04/04/2016 15:1...	new http://scholarsportal.inf...
/ALB_DV/doi-10-7939_D...	Alberta Survey, 2014A	04/04/2016 15:1...	new http://scholarsportal.inf...
/ALB_DV/doi-10-7939_D...	Alberta Survey, 2012A	04/04/2016 15:1...	new http://scholarsportal.inf...

Page 1 of 1

Displaying topics 1 - 42 of 42

<odesi> admin tool



17,567

Whoa! That's a big number, aren't you proud? Number of studies in
<odesi> after integration

Future Directions

Near future

Better graphing w/ weights
Sorting
DDI to Marc

Not so near future

Separate document
support

Proposed

Combine survey
variables
Add to Map



THANKS!

Any questions?

You can find us at

datagis@scholarsportal.info

