

## General concepts: DDI



**Irena Vipavc Brvar, ADP**

*SEEDS Kick-off meeting, Lausanne, 4. - 6. May 2015*



## How to describe our survey

What we learned so far:

If we want to use data at some point in the future, data need to be properly **documented**, **saved** in trusted place. Users need to be able to **find** and **access** it.

How do we achieve that. - > by using a standard

We would like surveys in our institutions to be describe in the same way. So every new colleague would know how to do it.

And possibly we would like to use such a standard that is used in similar organizations – interoperability between DA.

**DDI stands for Data Documentation initiative.** - > to establish a standard for technical documentation describing social science data.

# History

Idea was to produce metadata specification for the description of social science data resources.

- Initiated in **1994 (ICPSR)** / **XML** DTD already in **1997**

Contributors to the efforts of the DDI come from social science data archives and libraries in USA, Canada and EU and from major producers of statistical data (like the US Bureau of the Census, the US Bureau of Labour statistics, Statistics Canada and Health Canada)

- to replace the existing and widely used OSIRIS Codebook/data dictionary standard with a more modern and Web-aware specification.

- The first official version of the DDI specification (version **1.0**) was published in March **2000**.

- **2002 V2.0**
- **2008 V3.0**

(Ryssevik, 2001)



## 2 development lines

### **DDI-Codebook**

[DDI-Codebook](#) is a more light-weight version of the standard, intended primarily to document simple survey data. Originally DTD-based, DDI-C is now available as an XML Schema. The current version of **DDI-C** is 2.5.

### **DDI-Lifecycle**

Encompassing all of the DDI-Codebook specification and extending it, [DDI-Lifecycle](#) is designed to document and manage data across the entire life cycle, from conceptualization to data publication and analysis and beyond. Based on XML Schemas, DDI-Lifecycle is modular and extensible. Current **DDI-L** 3.2.

## BASIC STRUCTURE OF DDI 2.\*

- Section 1.0 - **Document Description** consists of bibliographic information that can be considered as the header whose elements uniquely describe the full contents of the compliant DDI file.
- Section 2.0 - **Study Description** consists of information about the data collection. This section includes information about who collected and who distributes the data, about the scope and coverage, sampling (if relevant), data collection methods and processing, citation requirements, etc.

## BASIC STRUCTURE OF DDI 2.\*

- Section 3.0 - **Data Files Description** provides information about the Data file(s).
- Section 4.0 - **Variable Description** provides a detailed description of variables, including (when relevant) the variable type, variable and value labels, literal questions, computation or imputation methods, instructions to interviewers, universe, descriptive statistics, etc.
- Section 5.0 - **Other Study Related Materials** allows for the inclusion of other materials related to the study such as questionnaires, user manuals, computer programs, interviewer manuals, maps, coding information, etc.

**Controlled vocabulary** (CESSDA topic classification, ELSST, DDI vocabulary)

## **Multilingual support**

- > ***CESSDA Catalogue***

Approximate number of elements in each specification

DDI 3.1 - 900

DDI 3.2 - 1150

DDI 2.1 - 400

DDI 2 Lite - 80

# PREPARING METADATA

Prepare a form in which researcher will insert information about the survey you need.



Gain clean data and other materials.



Prepare data and materials for long term preservation and distribution.



Prepare metadata description of the survey using information in the form (important – who are the authors (main and other), add project ID – funding – OpenAIRE compatible.) - Use tools // Possible export of question text, basic frequencies and descriptive statistics.



Distribute metadata (web, Nesstar etc.) - Make XML openly available – CESSDA catalogue // question bank

## Some data about Nesstar usage

Nesstar is currently run **by most archives in Europe**, and a reasonable number of data libraries in US/Canada.

Nesstar was originally **developed by and for archives**, and is designed to fit many important documentation and dissemination use-cases for data archives. Nesstar was also the first tool to **support DDI**, which is still a highly relevant standard for data documentation.

There are currently **> 130 instances of Nesstar Server worldwide**, from Vancouver to Taiwan and from South-Africa to Iceland.

In volume, the **International Household Survey Network** (<http://ihsn.org/home/>) is the most important Nesstar user. IHSN do not use Nesstar Server, but they use Nesstar Publisher as a documentation tool for statistical agencies in a large number of (developing) countries on all continents.

Nesstar also fully supports **multilingual metadata**, which makes it possible to document data in more than one language (without duplicating data).

Nesstar Server comes **with a set of APIs** that allow for third-party integration with data, metadata and functionality (e.g. tabulation and download operations) on the server.

Because of the APIs and the DDI support, the Nesstar platform is also very easy to repurpose for other services, e.g. the **CESSDA Portal** and the **DwB Data Discovery portal**.

Important/high profile users of Nesstar include:

[European Social Survey](#):



[UK Data Service](#)



[GESIS ZACAT](#)



It also supports aggregate data (cubes).  
[Norway's Institute of Public Health](#):

EUROPEAN ELECTION DATABASE

## Nesstar Publisher (*Located on desktop*)

**Nesstar Publisher** – a sophisticated authoring environment that can publish data from a variety of sources (including SPSS, SAS, Excel etc.). The tool includes a specialised metadata editor, data and metadata validation routines and metadata templates that provide standardisation and control.

Easy editing/creation and export of DDI documented datasets with XML experience needed.

Tools to compute/recode/label new, or existing, variables to be added to a dataset before publishing.

Tools to validate metadata and variables.

The ability to import and export data to the most common statistical formats, including delimited files.

The ability to include automatically generated frequency and summary statistics for each variable.

Multilingual - Arabic, Chinese, English, French, Portuguese, Russian and Spanish



# Nesstar Publisher

Projects:

- My Projects
  - studkr12-primer izpolnjene template
  - hrr11\_p1\_v1\_r1-full
    - rspStatment

**Naslov**

Ljudska prava u svijesti građana/ki Bosne i Hercegovine

**Podnaslov**

**Variables**

Number	Name	Label	Width	StartCol	EndCol	Record	Decimals
v1	m	Godina istraživanja	8	*	*	*	0
v2	m_1	Identifikacioni broj	4	*	*	*	0
v3	m_6	Region	1	*	*	*	0
v4	m_7	Tip naselja	1	*	*	*	0
v5	p_1	Ljudska prava	1	*	*	*	0
v6	p_2	Kako se šтите ljudska pra	1	*	*	*	0
v7	p_3_1	Koja ljudska prava vam p	3	*	*	*	0
v8	p_3_2	Koja ljudska prava vam p	3	*	*	*	0
v9	p_3_3	Koja ljudska prava vam p	3	*	*	*	0
v10	p_4_1	Pravo na život	2	*	*	*	0
v11	p_4_2	Pravo na slobodu i bezbj	2	*	*	*	0
v12	p_4_3	Pravo na jednakost pred	2	*	*	*	0
v13	p_4_4	Pravo na rad, izbor zapo	2	*	*	*	0
v14	p_4_5	Pravo na slobodu izražav	2	*	*	*	0
v15	p_4_6	Pravo na socijalno osigui	2	*	*	*	0

Category Hierarchy

- 1 - najvažnije
- 2 - 2
- 3 - 3
- 4 - 4
- 5 - 5
- 6 - 6
- 7 - najmanje važno

Value:  Label:

Category Text:

Level Name:

GeoMap URI:

**Variable information**

Data Type:

Measure:

Is Time Variable

Is Weight Variable

Min:  Max:  Decimals:

Implicit decimals

Missing data:

**Documentation**

Statistics Weights Documentation

- VarDoc
  - Pre-Question Text
  - Literal Question
  - Post-Question Text
  - Interviewer Instructions
  - Notes

**Pre-Question Text**

**Literal Question**

4.5 Rangirajte po važnosti 1 (najvažnije) 7 (najmanje važno) navedena ljudska prava po značaju: pravo na slobodu izražavanja, misli i vjeroispovjesti

# Nesstar Server (*Located on server*)

**Nesstar Server** - includes an SQL-based metadata management system, a data storage system, a powerful statistical engine as well as a flexible access control system.

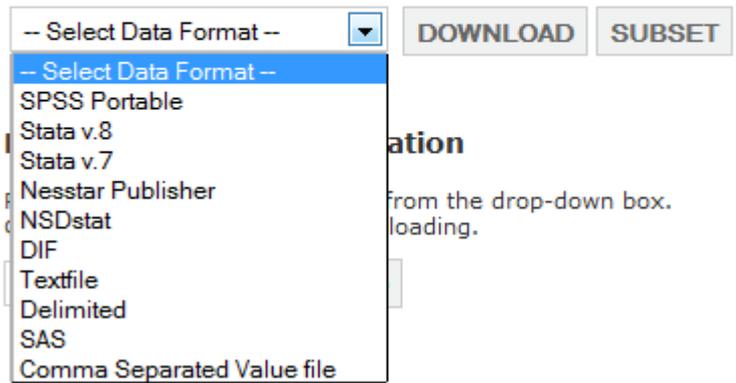
**Nesstar WebView** – totally customisable and configurable layer that presents the search, browse, display, analysis and retrieval options to the user. Able to seamlessly handle survey data, cubes and other resources.



Multiple crosstabulation and recoding

Regression and correlation analysis

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



- USDAS
  - Ljudska prava u svijesti građana/ki Bosne i Hercegovine
    - Metadata
      - Study Description
        - Bibliographic Citation
        - Study Scope
        - Methodology
        - Data Access
        - Other Study
      - Other Document
    - Variable Descriptions
      - Variable
      - Demografske v
      - Ostalo

**Depositor**

Name	Affiliation	Abbreviation
Centar za ljudska prava	Univerzitet u Sarajevu	HRC

**Dataset:** Ljudska prava u svijesti građana/ki Bosne i Hercegovine  
 Istraživanje javnog mnijenja o stanju ljudskih prava u Bosni i Hercegovini za 2011 godinu

## Variable p\_4\_2: Pravo na slobodu i bezbjednost

### LITERAL QUESTION

4.2 Rangirajte po važnosti 1 (najvažnije) 7 (najmanje važno) navedena ljudska prava po značaju: pravo na slobodu i bezbjednost

Values	Categories	N	
1	najvažnije	58	5.9%
2	2	543	55.4%
3	3	156	15.9%
4	4	107	10.9%
5	5	47	4.8%
6	6	43	4.4%
7	najmanje važno	26	2.7%
0		16	

### SUMMARY STATISTICS

Valid cases	980
Missing cases	16
Minimum	1.0
Maximum	7.0
This variable is numeric	

## Using Common Metadata for Harmonisation for Data Integration

**The CESSDA portal** is an example of integration of data in heterogeneous, autonomous resources (data archives) by using harmonised descriptive metadata represented in a common metadata standard, and using controlled vocabularies and code schemes.

Harmonisation of metadata is done by the DAs, and the harmonised metadata are made available in local servers for harvesting, and for presenting in the CESSDA portal.

<- Retaining the autonomy of the resources/DAs.

*More in [Deliverable 7.1 and 7.2-3](#)  
of DwB project*

## Events

- [EDDI](#) – yearly conference (since 2009) / also in the US
- [DDI workshops](#) in Castle Dagstuhl (since 2007)
- Presentations that are related to DDI at [IASSIST conferences](#)
- Trainings organized by CESSDA archives / CESSDA expert seminars

DDI Alliance [<http://www.ddialliance.org/>, 2. 5. 2015]

IHSN: Metadata Editor (Nesstar Publisher 4.0.9)

[<http://www.ihsn.org/home/software/ddi-metadata-editor>, 2.5.2015]

IHNS (2007): Quick Reference Guide for Data Archivists

[[http://www.ihsn.org/home/sites/default/files/resources/DDI\\_IHNS\\_Checklist\\_OD\\_06152007.pdf](http://www.ihsn.org/home/sites/default/files/resources/DDI_IHNS_Checklist_OD_06152007.pdf), 2.5.2015]

Ryssevik, J. (2001). The Data Documentation Initiative (DDI) metadata specification. Paper prepared for MetaNet 2001, Voorburg, Netherlands.

[[http://www.ddialliance.org/sites/default/files/ryssevik\\_0.pdf](http://www.ddialliance.org/sites/default/files/ryssevik_0.pdf), 2. 5. 2015]

Martinez, L. (2008): The Data Documentation Initiative (DDI) and Institutional Repositories

[[http://www.disc-uk.org/docs/DDI\\_and\\_IRs.pdf](http://www.disc-uk.org/docs/DDI_and_IRs.pdf), 2.5.2015]

