

ČSDA

Czech Social Science Data Archive

Jilská 1, 110 00 Praha 1, Czech Republic; <http://archiv.soc.cas.cz>

Martin Vávra
Tomáš Čížek

Data Acquisition in Practice in CSDA

**CES2012: Data Acquisition and License Agreements,
Tampere, Finland**

Data Acquisition in CSDA & Licensing

- ▶ About CSDA
- ▶ Data sources in the Czech Republic
- ▶ Historical datasets
- ▶ Practice of data acquisition
- ▶ Licensing

- **The Czech Social Science Data Archive (CSDA)** at the **Institute of Sociology, Academy of Sciences of the Czech Republic (ASCR)**, serves as the national focal point providing open access to electronic files of primary data from research projects.
- A publicly accessible social science data archive was established at the Institute of Sociology, Academy of Sciences of the Czech Republic in 1998 under the title Sociological Data Archive (SDA). In 1999, it became an independent research department of the Institute. In 2011, the name of the department was changed to Czech Social Science Data Archive (CSDA).

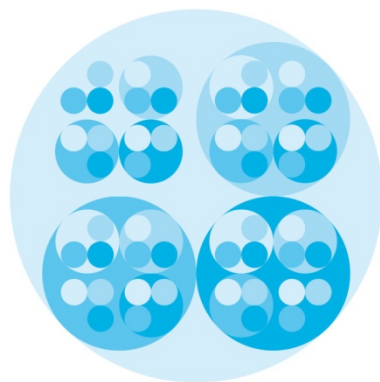
Acquisition from all Social Sciences



SDA Sociological Data Archive (1998 - 2011)



Czech Social Science Data Archive



ČSDA

Data sources in the Czech Republic



- Primary source of our data is our home institute.
- Institute of Sociology is leading sociology department in the Czech Republic – majority of sociological surveys comes from it. (including international projects like ISSP, ESS)
- Universities with sociological programs
- Government institutions
- Official statistics !
- Commercial companies
- Same goes for other social sciences:

Central register of R&D projects



- Covers research projects from 1993 (Dissolution of Czechoslovakia)
- Each science have unique code
- 788 research projects in the field of sociology and demography
- More than 250 include some form of empirical research
- 335 projects of political science, 249 of psychology,...

Central register of R&D projects

Sociology and Demography	Count	%
Empirical research projects	257	33
Other projects without empirical research	486	62
Unclear	45	5
Total	788	100

Historical data

- 🌐 One possible (and fruitful) target of our data acquisition enterprise are survey data from period 1946 -1990
- 🌐 Big part of primary data sets are lost but we are trying to utilize reports on those old surveys

Data from 40'

In 1946 Institute for public opinion polls was established

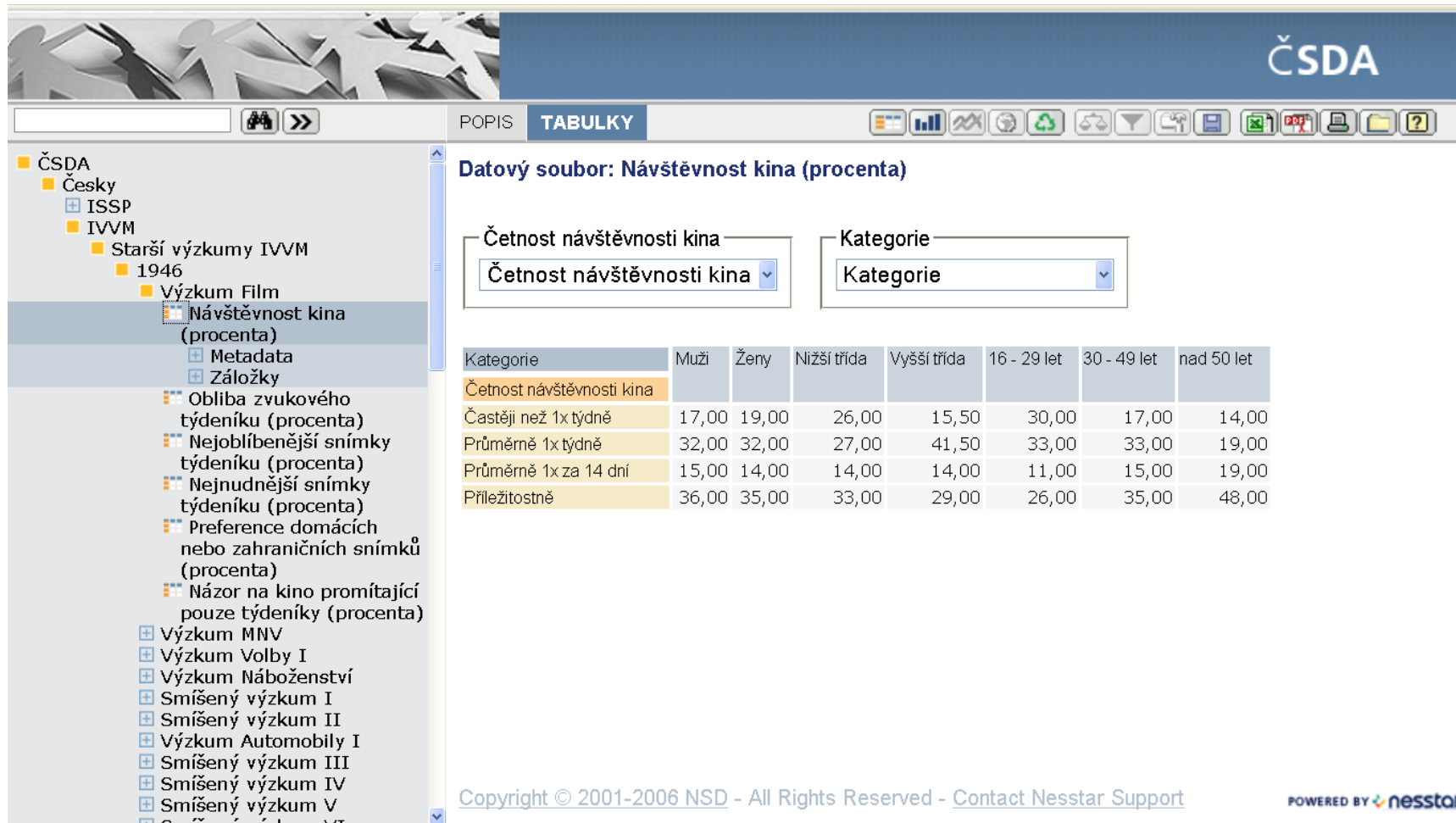
Primary data are lost, but we have results in form of frequencies and crosstabulations and we are publishing them via Nesstar as „Nesstar cubes“

10. Je ilustřý kojenec odolnější vůči nemocem! Ano 1
 Ne 2
 Nevím 3

B. 52
 Analýza
 Osobní údaje
 v říjnu

Základ	Úhrn	Z úhrnu připadá na odpovědi:						Poznámka
		č. 1	č. 2	č. 3	č. 4	č. 5	č. 9	
Všichni	1000	178	548	269			5	
		18	55	27				
Muži	459	90	226	142			1	
		20	69	31				
Ženy	541	88	322	127			4	
		16	60	24				
18—29	223	37	129	57				
		17	58	25				
30—49	466	43	263	129			1	
		16	56	28				
50 —	311	68	156	83			4	
		22	51	27				
Dělníci <i>muži</i>	231	47	107	77				
		20	46	34				
Úředníci a zřizenci <i>muži</i>	79	12	51	16				
		15	65	20				
Zemědělci <i>muži</i>	88	19	38	30			1	
		22	44	34				
Ostatní samostatní <i>muži</i>	61	12	30	19				
		20	49	31				
<i>ženy</i>	233	38	133	58			4	
		14	54	25				


Historical data turned into Nesstar Cubes



The screenshot displays the Nesstar Cubes interface. The top navigation bar includes the ČSDA logo and a 'TABULKY' (Tables) tab. The left sidebar shows a hierarchical tree structure of data sources, with 'Návštěvnost kina (procenta)' selected. The main content area is titled 'Datový soubor: Návštěvnost kina (procenta)'. It features two dropdown menus for filtering: 'Četnost návštěvnosti kina' (set to 'Četnost návštěvnosti kina') and 'Kategorie' (set to 'Kategorie'). Below these is a data table with columns for 'Kategorie', 'Muži', 'Ženy', 'Nižší třída', 'Vyšší třída', '16 - 29 let', '30 - 49 let', and 'nad 50 let'. The table contains five rows of data representing different attendance frequencies.

Kategorie	Muži	Ženy	Nižší třída	Vyšší třída	16 - 29 let	30 - 49 let	nad 50 let
Četnost návštěvnosti kina							
Častěji než 1x týdně	17,00	19,00	26,00	15,50	30,00	17,00	14,00
Průměrně 1x týdně	32,00	32,00	27,00	41,50	33,00	33,00	19,00
Průměrně 1x za 14 dní	15,00	14,00	14,00	14,00	11,00	15,00	19,00
Příležitostně	36,00	35,00	33,00	29,00	26,00	35,00	48,00

Copyright © 2001-2006 NSD - All Rights Reserved - [Contact Nesstar Support](#)

POWERED BY 

Data from 60' - 80'

- The same process is used for archiving most of the data produced during this period
- Sometimes there are primary data available but they are saved on out-of-date media
- In this case finishing acquisition means looking for some hardware able to read old magnetic tapes

This version...



Or even this version....



Sometimes we are successful



- 🌐 Acquisition is quite effective among researchers at Institute and departments of sociology at universities – it is partly because of our personal relations, partly because we have something to offer them (e. g. lectures for students on using Nesstar)
- 🌐 Moreover we are „well known“ among these people already – they know our activities and mostly are willing to collaborate

- 🌐 Much more difficult is situation with out-of-sociology researchers and/or with researchers from non-university institutions. They more frequently:
 - ▶ understand data as their „property“
 - ▶ do not trust in quality of their data and are afraid that archiving data and making them available for scientific community can question their reputation
 - ▶ did not perceive archiving as beneficial for them

Acquisition of qualitative data

- 👤 Very difficult
- 👤 Missing informed consents
- 👤 Each license agreement is unique

General principles

- 🌐 archived data remain the property of the depositor
- 🌐 depositor can set specific conditions for data availability
- 🌐 Archive gives an undertakings:
 - ▶ to provide data only for scientific purposes
 - ▶ to mention depositor and author of the data set in documentation
 - ▶ to commit users to quoting author of the data
 - ▶ to keep data undamaged and migrate data sets to up to date data formats

Thank You for your
attention !

