



**ARHIV  
DRUŽBOSLOVNIH  
PODATKOV**

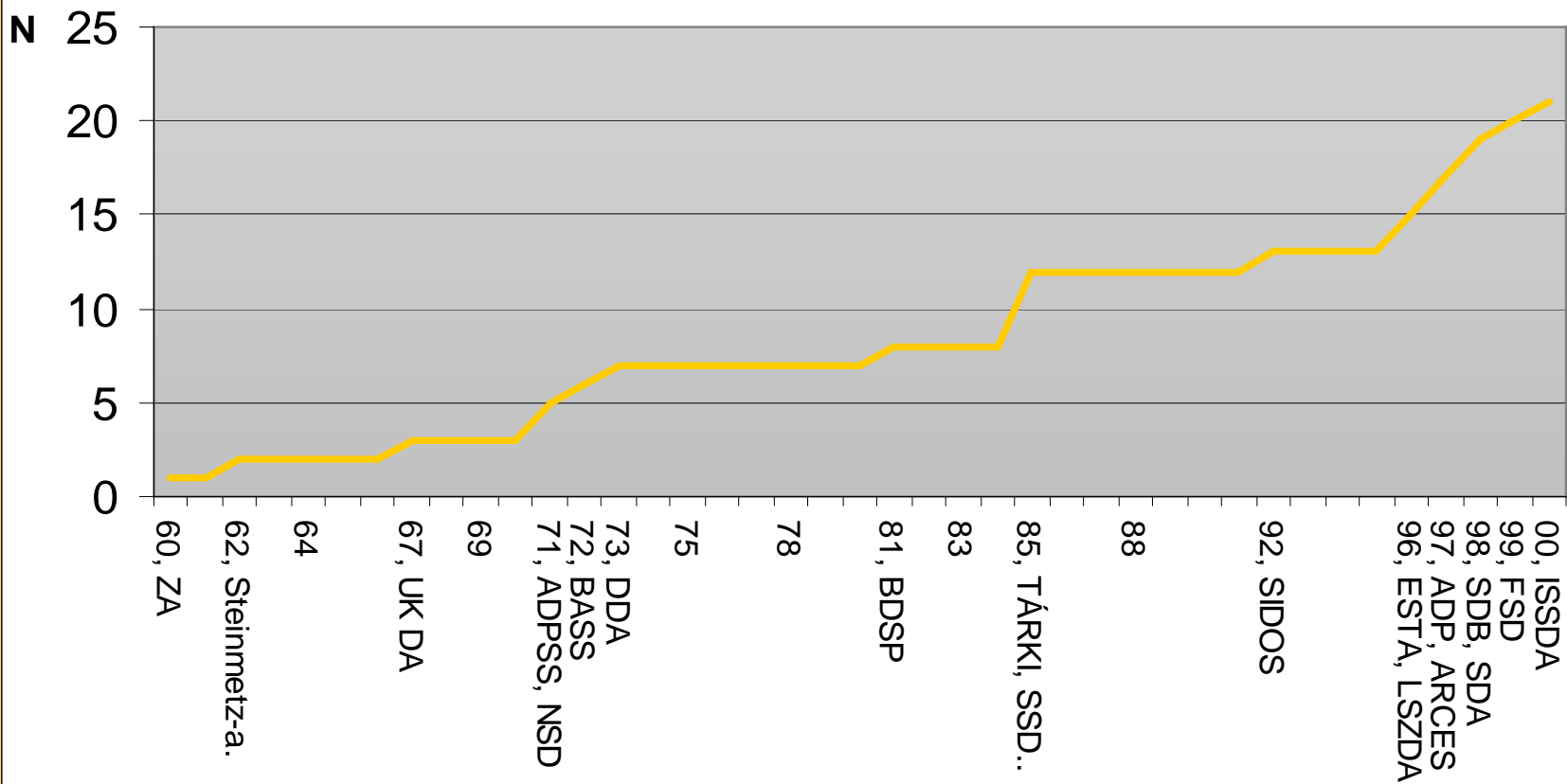
## **How to take advantage of existing information about studies when populating a new data archive: the ADP's experience**

**Janez Štebe and Irena Vipavc**  
**ADP - Social Science Data Archive, Slovenia**

IASSIST/IFDO conference 2001, 14 –19 May, Amsterdam

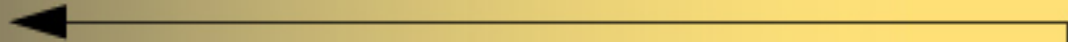


## Emergence of national data archives over time (\*1)





**ARHIV  
DRUŽBOSLOVNIH  
PODATKOV**



**Situation at the end of nineties: a new social science data archive is one among many similar institutions in a field of science information and research infrastructure.**

**It has to define its position by searching similarities and differences compared to others.**

**The ADP is defined geographically, covering the material of a particular state or regional origin, and thematically, regarding the content area that it covers, and by sort of material that it collect and disseminate.**

**The ADP is *Slovene national (geographical) social science (thematic) computer – readable data archive (sort of material)***



**ARHIV  
DRUŽBOSLOVNIH  
PODATKOV**



## **Comparison of ADP and other related institutions and information providers**

### **Institution in comparison:**

**other national social science data archives (CESSDA, IFDO)**

### **Differences:**

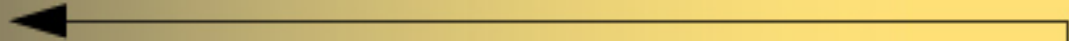
**geographic scope; thematic emphasis;**

### **Similarities:**

**same sort of material, services offered; same data sets in case of multinational studies**



**ARHIV  
DRUŽBOSLOVNIH  
PODATKOV**



**Institution in comparison:**

**specialised thematic data archives, e.g. single study archive  
(RIS, ISJP, ISSP Home Page, ISSP at ZA, NDB, CSES, FFS,...)**

**Differences:**

**emphasis of service offered – more in depth, specialised offer, e.g. bibliography,  
original documents, addresses of participating countries**

**Similarities:**

**same primary unit of material – data set; take over part of service of an archive  
and enrich that in direction of more intensive study processing of single study**



**ARHIV  
DRUŽBOSLOVNIH  
PODATKOV**



**Institution in comparison:**

**COBBIS (Slovene Cooperative Online Bibliographic System and Services)**

**Differences:**

**sort of material – bibliographic references on publications stored in libraries in Slovenia; Slovene Bibliography**

**Similarities:**

**it is connected on same study or data set or author; services offered is distribution and information about material stored**



**ARHIV  
DRUŽBOSLOVNIH  
PODATKOV**



**Institution in comparison:**

**SICRIS (Slovenian Current Research Information System)**

**Differences:**

**broad level of processing, emphasis is on content and not on methodology and data,  
variables description**

**Similarities:**

**overlapping in higher level unit: same study, project, author;  
science information service**



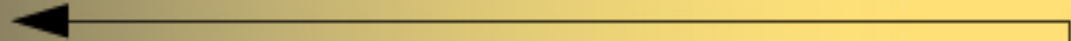
Information technology makes possible to transgress the boundaries of physical location.

- the information providers are more transparent;
- they grow in number;
- interconnections between single units of material stored in different locations are easier;
- linking of services and offering is better

Motto 1:

***"Obviously, integration and diversity (...) offer almost unlimited chances for creativity of research and new demands on the social science infrastructure. This challenging world continuously demands vigorous scrutiny of priorities, procedures and outcomes."***

**(Mochmann, 1998)**



## **Purpose of the presentation**

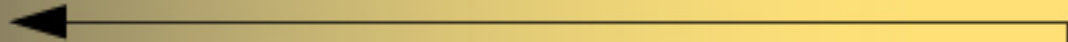
**How to undertake the acquisition and archiving of data to be maximally effective?  
In shortest time and with minimal cost to provide maximal number of relevant  
information's and product to users!**

**Which orientations can we extract from overview of**

- existing electronic information about studies,**
- possibilities offered by a new standard DDI DTD XML Codebook,**

**that is to say by exploring the possibilities of**

***"Collaborative Working in the Social Science Cyber Space"***



## Users expectations and image of the Archive

### Motto 2:

*"Data-archives have a function as national repositories for datasets. They have presented themselves as such and they are seen as such."* (Mochmann and Guchteneire, 1988)

### Scheme: Congruity of Data archive offer and Requirements of users

Profile of the arcive		Requirements of users	
unexploited Archive capacities	Archive capacities used, fulfilled users expectations	not realized demands	

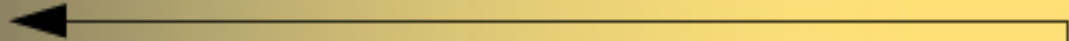


**New archives lag behind in acquisitions compared to the well-established archives.**

**How to direct the processing of studies to achieve goals?**

**Dilemmas:**

- **extensive processing of large number of cases or intensive processing of small selection**
- **requirements of users (Lievesley, 1999):**
  - **one stop shopping**
  - **timelines**



**Solution to our problem is Cost effective strategy of use of resources,**

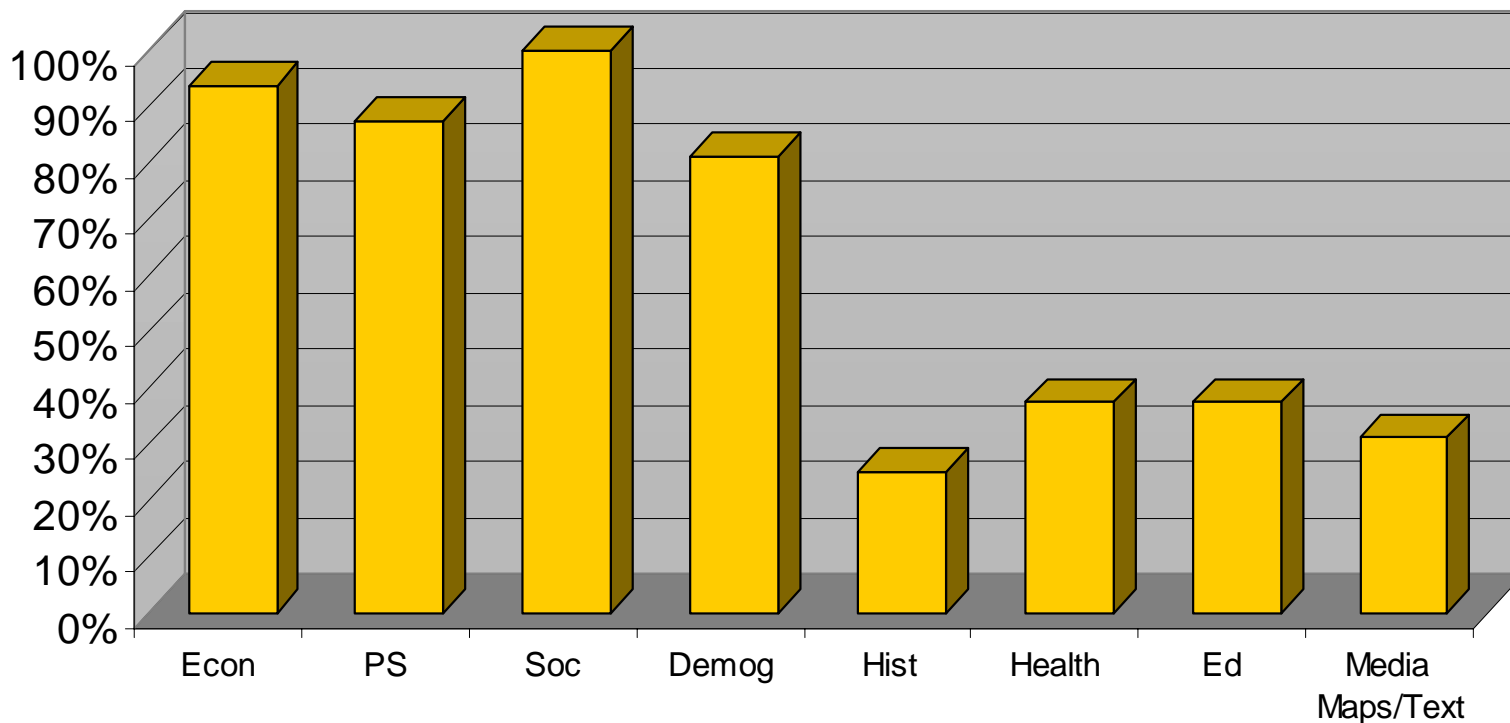
**To *set priorities* and *make a selection* in phase of aquisition and adapt intensity of processing taking into account**

- A. thematic, content area in accordance with the profile of the Archive**
- B. estimated "future usability" and relevance of dataset for secondary analysis taking into balance**
- C. cost and time of data set preparation, considering also possibilities of**
- D. use of existing information and other available resources through cooperation**



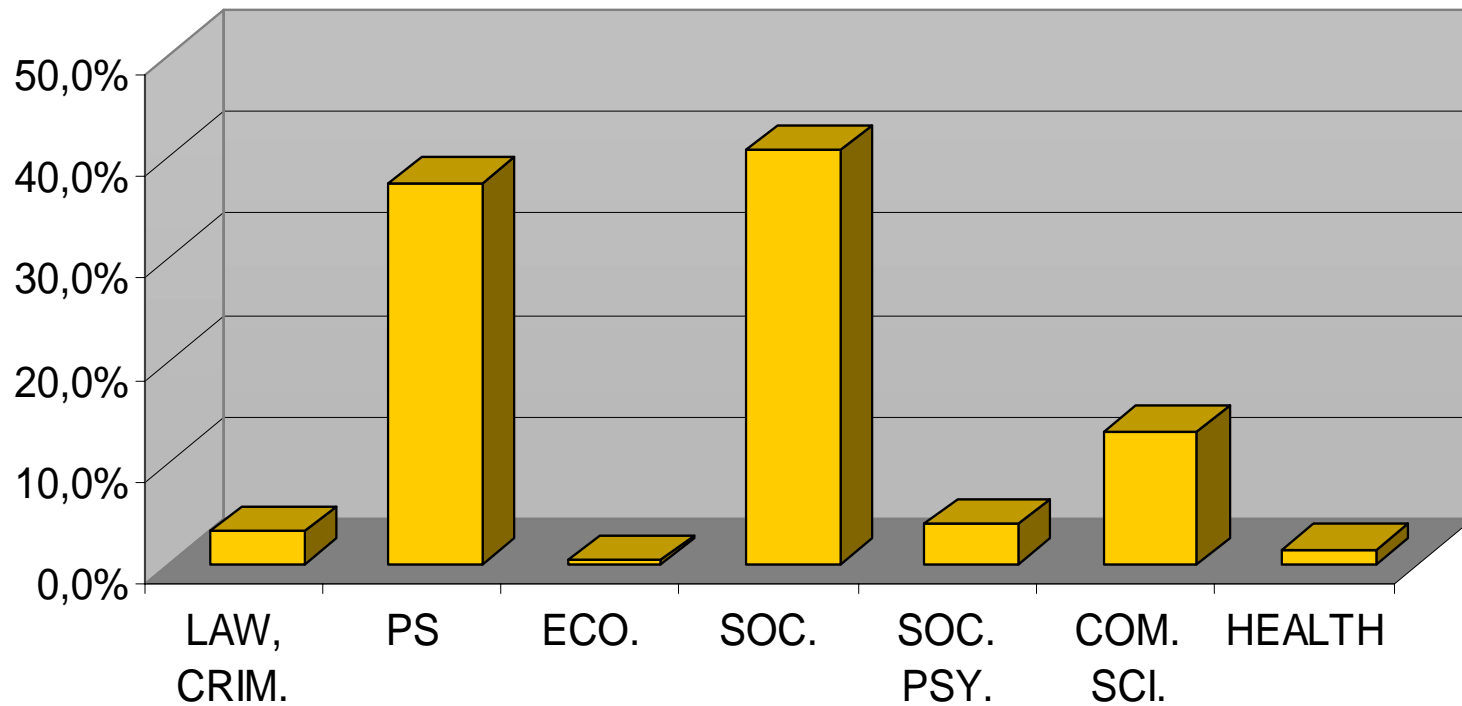
**A) thematic, content area profile of the Archive**

**Subject coverage in Social Science and Economic Data Archives  
(Source: Guy and Haley, 1995; N = 17)**





**Studies in ADP according to subject classification (CERIF)**  
(Source: ADP studies holding, April 2001; N = 283) (\*2)





## Thematic profile of the Archive

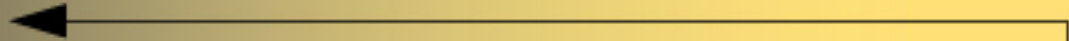
Archive profile (specialization)	Users expectations
national / thematic specialisation Taking advantage of close knowledge of material to provide more accurate information.	one stop shopping Excluded are: <ul style="list-style-type: none"><li>• other national</li><li>• other thematic areas</li></ul>

### Some strategies to achieve users satisfaction

- use of standards (DDI DTD) and related services of virtual data catalogues
- provide basic study description of units falling on an edge of archive profile
- encourage the emergence of similar institutions for related areas or find additional founding to enable broader coverage



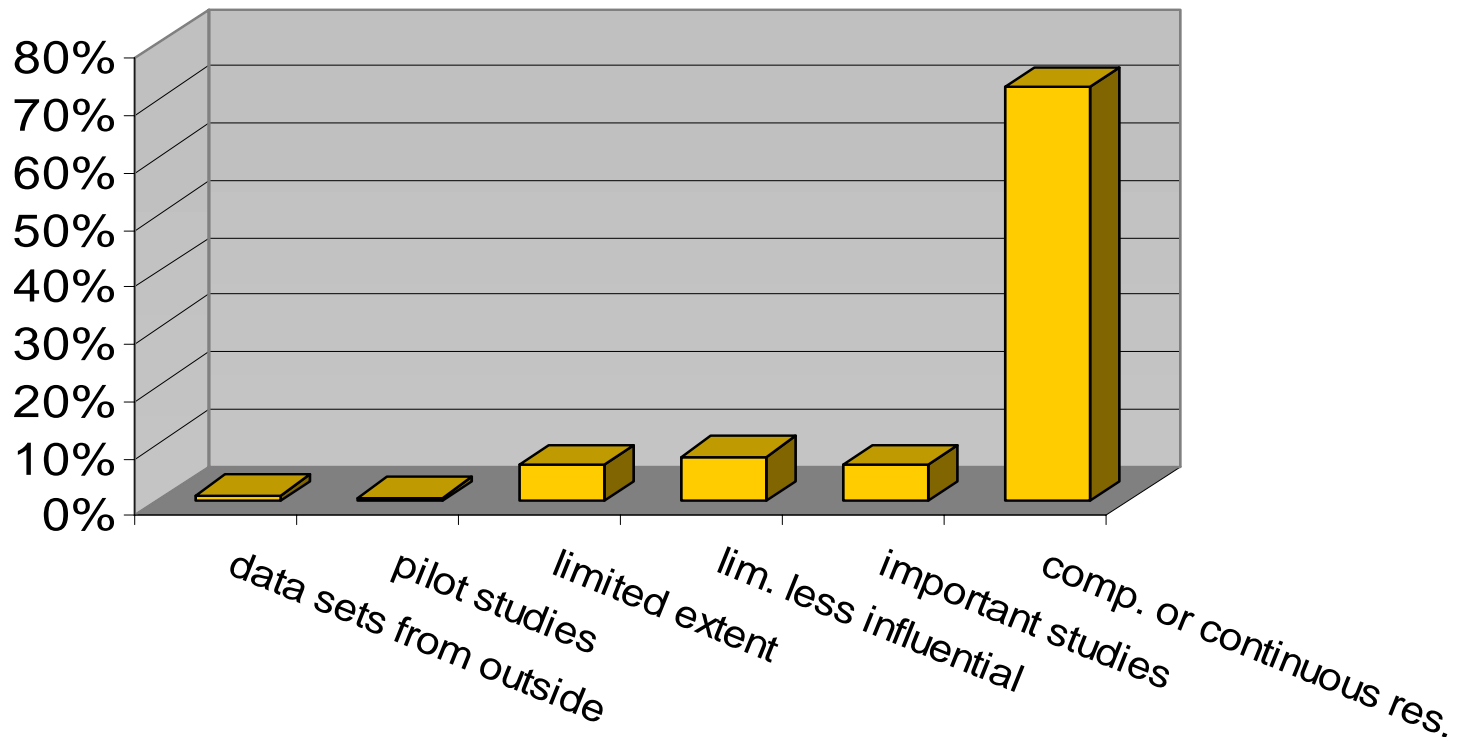
**ARHIV  
DRUŽBOSLOVNIH  
PODATKOV**



**B) estimated "future usability" and relevance or dataset for secondary analysis, restrictions on usage of data, proper documentation available  
(See: Mochmann and Guchteneire, 1988; Lievesley, 1999)**



**Studies in ADP clasified according to relevance**  
(Source: ADP studies holding, April 2001; N = 283) (\*3)





Archive profile	Users expectations
long-term preservation of data highly relevant scientific purposes, freely available to users	different users and purposes on one place  Excluded are: <ul style="list-style-type: none"><li>• non-typical or non-central purpose</li><li>• less important studies</li><li>• studies not yet deposited in an Archive</li></ul>

### Some strategies to achieve users satisfaction

- tell the users what is the primary purpose of Archive and who are the "target" users, education of users
- provide basic study description of units low on relevance criteria, or not yet deposited in an Archive by pointing to publications and other additional sources of information



### **C) cost and time minimisation**

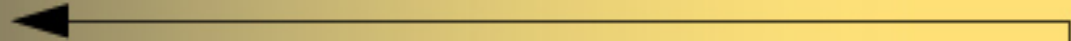
Archive profile	Users expectations
<ul style="list-style-type: none"><li>• minimise time and cost on an unit of information</li><li>• but keep up with the mission of an Archive of an intensive treatment of most important studies, with high standards of quality of processing</li><li>• find out where an Archive with its specialization can cover the niches in the current information offer available elsewhere with the goal of adding value to the data sets</li></ul>	

#### **Some strategies to achieve users satisfaction:**

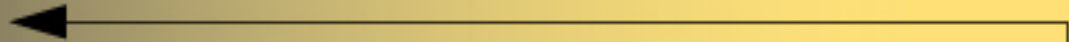
- **maximise use of existing information to save on time and money**
- **provide basic study description immediately, before the end of processing**
- **additional founding to enable archiving of the historical legacy of social science studies in a country**



**ARHIV  
DRUŽBOSLOVNIH  
PODATKOV**

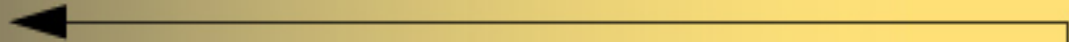


**Examples from the Slovenian Social Science Data Archive (ADP)  
study processing as an illustration of the utilisation of existing  
electronic information available elsewhere**



## Related institutions

- Other national archives
- Specialised single study archives
- SICRIS - **S**loven**I**an **C**urrent **R**esearch **I**nformation **S**ystem
- COBBIS - **C**ooperative **O**nline **B**ibliographic **S**ystem and **S**ervices



## Units of study and data description

- content and methodology description
- author, institution
- publications
- summary data overview
- raw data, codebook



<i>Units of study and data description</i>	<i>Related institutions</i>			
	Other national archives	Specialised single study archives	SICRIS	COBBIS
<b>content and methodology description</b>	☹  (WVS_95)	☹  (RISZAV98)	☹  (MLA98)	
<b>author, institution</b>			☹  (MLA98)	☹  (MLA98)
<b>publications</b>		☹  (FFS97)		☹  (RISZAV98)
<b>summary data overview</b>	☹  (NESSTAR)	☹  (NDB)		☹  (RISZAV98)
<b>raw data, codebook</b>	☹  (ISJP)	☹  (CSES)		



- CESSDA - <http://www.nsd.uib.no/cessda/>
- COBBIS - Slovene Cooperative Online Bibliographic System and Services  
<http://www.izum.si/COBISS/>
- CSES - The Comparative Study of Electoral System <http://www.umich.edu/~nes/cses/cses.htm>
- FFS - Fertility and Family Surveys <http://www.unece.org/ead/pau/ffs/>
- FFS97 - [http://rcul.uni-lj.si/~fd\\_adp/opisi/FFS97-en.XML](http://rcul.uni-lj.si/~fd_adp/opisi/FFS97-en.XML)
- IFDO - <http://www.ifdo.org/>
- ISJP - International Social Justice Project <http://www.isjp.de/>
- ISSP - International Social Survey Programme <http://www.issp.org/>
- MLA98 - [http://rcul.uni-lj.si/~fd\\_adp/opisi/MLA98.XML](http://rcul.uni-lj.si/~fd_adp/opisi/MLA98.XML)
- NDB - New democracies barometer <http://www.cspp.strath.ac.uk/>
- RIS - Research on Internet in Slovenia <http://www.ris.org/indexuk.html>
- RISZAV98 - [http://rcul.uni-lj.si/~fd\\_adp/opisi/RISZAV98.XML](http://rcul.uni-lj.si/~fd_adp/opisi/RISZAV98.XML)
- SICRIS - Slovenian Current Research Information System <http://sicris.izum.si/>
- VWS\_95 - [http://rcul.uni-lj.si/~fd\\_adp/opisi/WVS\\_95.XML](http://rcul.uni-lj.si/~fd_adp/opisi/WVS_95.XML)



## **(1\*) List of data archives with the year of establishment**

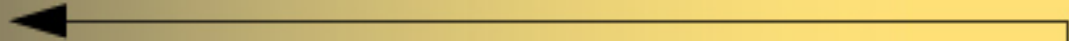
Emergence of national data archives over time

Year Name

1960	ZA – Central Archive for Empirical Social Research, University Cologne
1962	Het Steinmetz-archieff – the Netherlands
1967	UK DA - The UK Data Archive
1971	ADPSS – Data Archive for Social Science – Italy
1971	NSD - Norwegian Social Science Data Services
1972	BASS – Belgian Archives for the Social Sciences
1973	DDA - The Danish Data Archives
1981	BDSP - France - Banque de Données Socio-Politiques
1985	TÁRKI - Data Archive and Information Technology Department - Hungary
1985	WISDOM - Das Wiener Institut für sozialwissenschaftliche Dokumentation und Methodik, Austria
1985	SSD - Swedish Social Science Data Service
1985?	ISDC - Israel Social Sciences Data Center ?
1992	SIDOS - Swiss <u>Data Archive</u> for the Social Sciences
1996	ESTA - Estonian Social Science Data Archiv
1996	LSZDA – Latvian Social Science Data Archive



**ARHIV  
DRUŽBOSLOVNIH  
PODATKOV**



1997	ADP - Slovene Social Science Data Archive
1997	ARCES (Social Studies Archive), Spain
1998	SDB – Greek, Social Data Bank
1998	SDA – Sociological Data Archive Czech Republic
1999	FSD - The Finnish Social Science Data Archive
2000	ISSDA – Irish Social Science Data Archive



**(\*2) Studies in ADP according to subject classification (CERIF)**

(Source: ADP studies holding, April 2001; N = 283)

	Freq.	Valid Percent	Cumulative Percent
LAW, CRIMINOLOGY	9	3,2%	3,180212
POLITICAL SCIENCES, POLEMOLOGY	106	37,5%	40,63604
ECONOMICS	1	0,4%	40,9894
SOCIOLOGY	115	40,6%	81,62544
SOCIAL PSYCHOLOGY	11	3,9%	85,51237
COMUNICATION SCIENCES	37	13,1%	98,58657
HEALTH	4	1,4%	100
Total	283	100	
Missing System	1		
Total	284		



**(\*3) Studies in ADP clasified according to relevance  
(Source: ADP studies holding, April 2001; N = 283)**

	Freq.	Valid Percent
data sets outside the scope of the archive; problems covered elsewhere; occasional studies of limited extent and of low quality	3	1,1%
pilot studies that create or improve the instrument	2	0,71%
studies of limited theoretical or practical conceptual extent	18	6,36%
less influential studies that permits theoretical generalisations or relates on a practical problem	22	7,77%
theoretically or practically important studies that fill the research gap	18	6,36%
comparative or continuous research, influential populations, with methodological excellence	205	72,43%
Total	283	100%