

Data Discovery: Vyhledávání výzkumných dat pro vědeckou analýzu

Prezentace pro Open Science Week 2024

Johana Chylíková
Johana.chylikova@soc.cas.cz
ČSDA

Data Discovery: Vyhledávání výzkumných dat pro vědeckou analýzu

TLDR: Ukážeme si zdroje různých typů dat

- Studie v sociálních vědách různého typu (různé typy dat z různých oblastí)
- Ne vždy máme prostředky na sběr vlastních dat
- Možnost využít existující data

Jaké typy dat můžeme vyhledávat?



Data z výběrových šetření (survey data)

Kvantitativní data získaná prostřednictvím standardizovaných dotazníků na vzorcích populace

- Průřezová, longitudinální
- Soudobá, získaná v posledních cca 20 letech, ale i historická

Survey data k nalezení ve všech sociálně-vědních datových archivech CESSDA (Evropa)

- V CESSDA archivech důvěryhodná, metodologicky korektně získaná survey data
- Zpravidla získaná v rámci sociálněvědních akademických projektů

Velké archivy jako GESIS, UK Data Service, Nizozemský archiv DANS

Středně velké archivy jako Slovinský ADP či český ČSDA

Přehled všech CESSDA archivů a odkazy na jejich katalogy na webu <https://www.cessda.eu/>

CESSDA Data Catalogue

"One stop shop", vyhledávání v datech CESSDA archivů

(pozor, nejsou tam všechna data z CESSDA archivů, proto, když např. vyhledáváte česká data, jděte rovnou do katalogu ČSDA)

- Deskripce více 40 000 datových souborů z více než 20 evropských zemích
- Deskripce obsahují link na soubor v "domovském archivu"
- Vyhledávání podle země, let sběru, klíčových slov a tématu

<https://datacatalogue.cessda.eu/>



ICPSR

- Největší sociálněvědní datový archiv na světě, mezinárodní se sídlem v USA
- Největší sbírka souborů dat z výběrových šetření
- <https://www.icpsr.umich.edu/web/pages/>

Find Data

Search
[view all](#)

[search tips](#) ▾

Browse




[Topics / Series / Thematic data collections / Data-related publications](#)

- [Studies with online analysis](#)
- [Self-published data, including replication datasets](#)
- [Studies with learning guides](#)


New/Updated Data Releases:

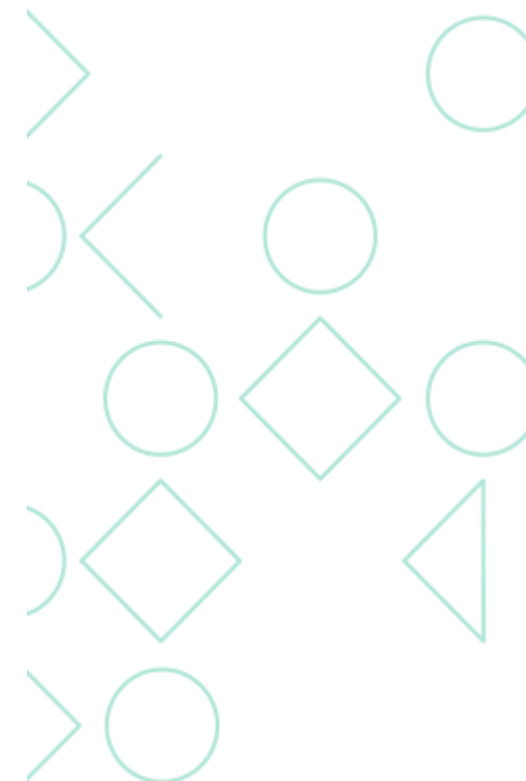
- [In the last week](#)
- [In the last month](#)
- [In the last quarter](#)

Statistics

-  [20 711 studies](#)
-  [6 593 772 variables](#)
-  [115 977 publications](#)

Most Popular Search Terms



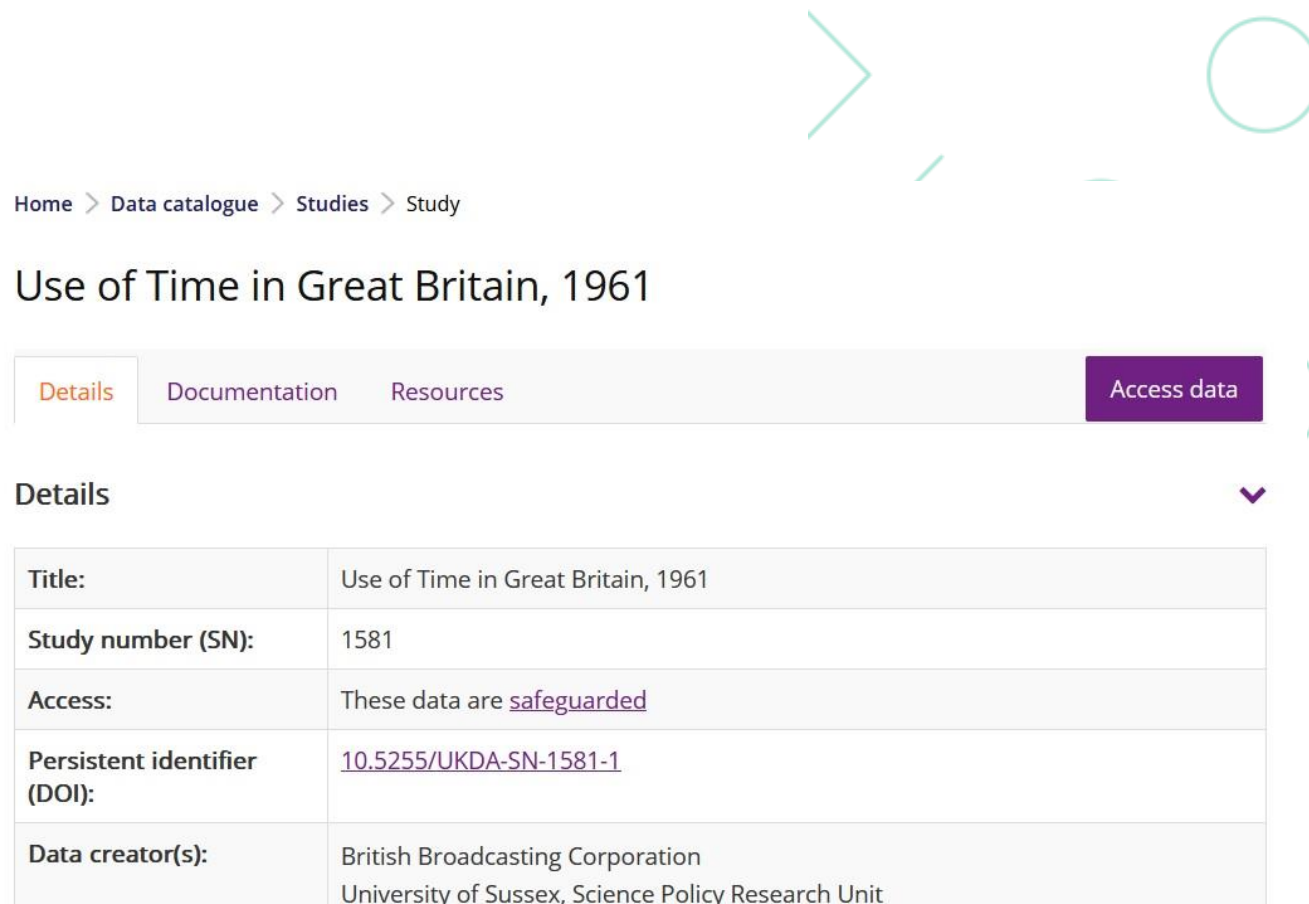


Historická survey data

V ČSDA lze najít survey data z 90. let 20. stol.

v Evropských a světových archivech i starší

- Např. UK Data Service
- Survey on time use z roku 1961
- Trávení času Britů na začátku šedesátých let
- <https://beta.ukdataservice.ac.uk/datacatalogue/studies/study?id=1581>



Home > Data catalogue > Studies > Study

Use of Time in Great Britain, 1961

Details Documentation Resources [Access data](#)

Details

Title:	Use of Time in Great Britain, 1961
Study number (SN):	1581
Access:	These data are safeguarded
Persistent identifier (DOI):	10.5255/UKDA-SN-1581-1
Data creator(s):	British Broadcasting Corporation University of Sussex, Science Policy Research Unit

Administrativní data

Statistiky, ukazatele sbírané státem za účelem správy

I v datových archivech (ale lze hledat i v databázích organizací, např. OSN)

Soudobá

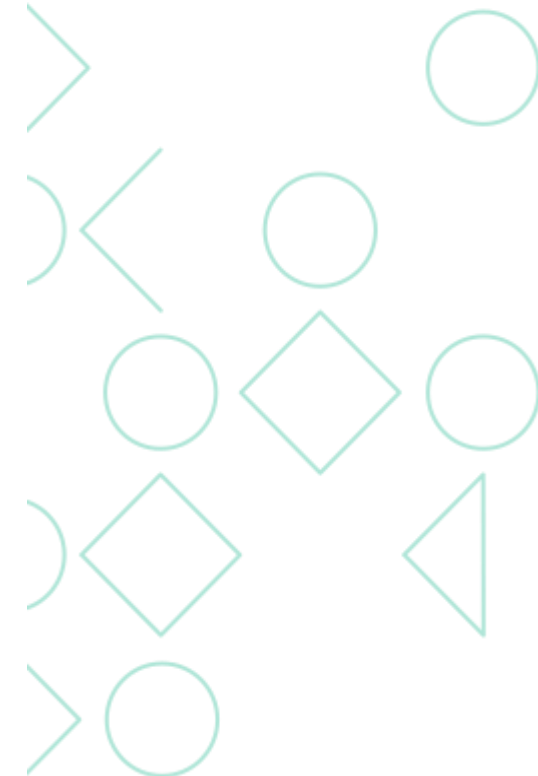
Data z cenzů (sčítání lidu) a ekonomická data

Příklad dat:

- ICPR, US Census data
- UKDS United Kingdom Census data
- UKDS Business microdata

Historická

Např. Databáze židovské komunity v Liverpoolu (1740-1881), UK Data Service,



Kvalitativní výzkumná data

- Především texty (přepisy rozhovorů, focus group)
- Soudobá i historická
- Zpravidla omezený přístup, anonymizována, ochrana osobních údajů

Příklad dat:

UK Data Service

Girlhood and Later Life: Girls Growing Up in Britain 1954-1976 and the Implications for Later-Life Experience and Identity, 1939-2020

Hlubkové rozhovory



Data ze sociálních sítí

Příklad:

- GESIS Web Data Projekt
- Databáze: “Digital Behavioral Data” v katalogu GESIS



GESIS Leibniz-Institut
für Sozialwissenschaften

Příklad dat:

Invasion Ukraine, příspěvky na Twitteru z období 17. 2. - 3. 3. 2022, sebrané prostřednictvím Twitter Streaming API, 2,3 miliony uživatelů, tweety v AJ

https://search.gesis.org/research_data/SDN-10.7802-2555

Další zdroje pro vyhledávání dat

Registry datových repozitářů

- Re3data.org
 - přes 2000 datových repozitářů ze všech vědeckých oborů
 - vyhledávání podle kritérií (předmět, země atp.)
- OpenAIRE Explorer (OpenAIRE)
- OpenDOAR (Jisc), the Directory of Academic Open Access Repositories
- FAIRsharing

Datový vyhledávač

- Google Dataset Search



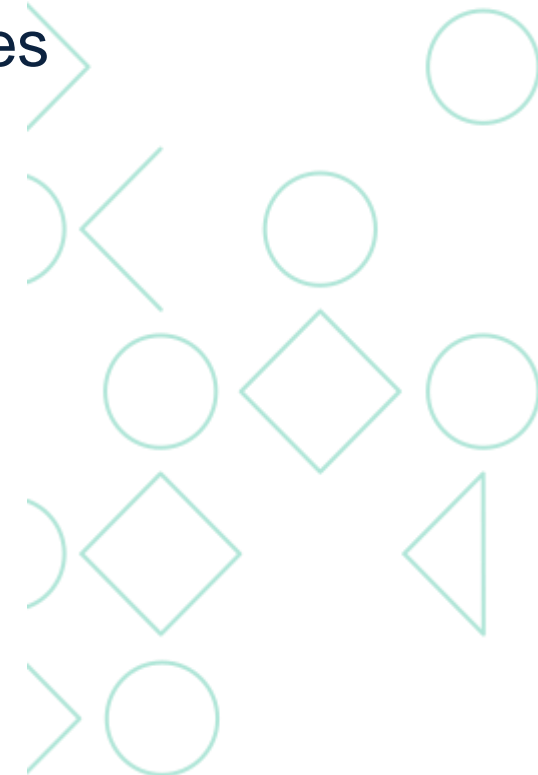
Další zdroje pro vyhledávání dat

Datový časopis

- Research Data Journal for the Humanities and Social Sciences
- Scientific data (Nature, n.d.)

Články informujících o datasetech, které vytvořili jejich autoři

- Historická, administrativní, demografická, ekonomická data
- Informace o nástrojích k extrakci, sběru dat



Další významné zdroje dat:

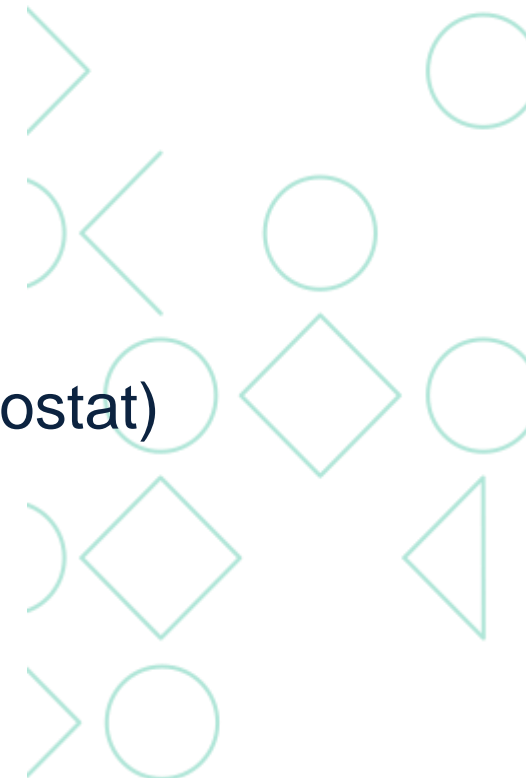
Mezinárodní komparativní projekty

Programy sociologického a socio-ekonomického výzkumu

Data k dispozici prostřednictvím webů projektů či Eurostatu

Kromě EU-SILC volný přístup

- EU-SILC; EU Statistics on Income and Living Conditions (Eurostat)
- Eurobarometer
- European Social Survey (ESS)
- International Social Survey Programme ISSP
- SHARE – Survey of Health, Ageing and Retirement in Europe



Děkuji za pozornost

Johana.chylikova@soc.cas.cz

