

Management dat ve výzkumné praxi

Workshop: Open science a sdílení sociálněvědních dat v praxi: Data management plan, ochrana respondentů a legální aspekty sdílení dat

Johana Chylíková, Ph.D.

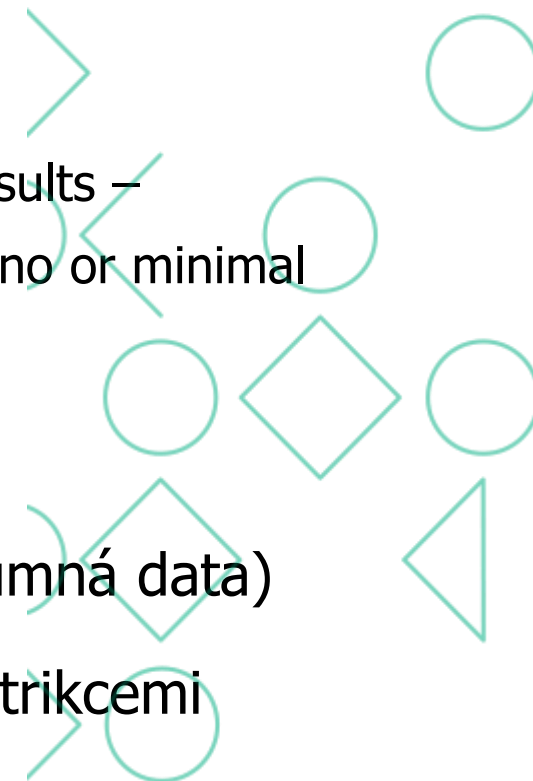
Český sociálněvědní datový archiv
Sociologický ústav AV ČR, v.v.i.

Sdílení dat v sociálních vědách

- Návaznost na koncept Open Science
- Open Science: "to make the primary outputs of publicly funded research results – publications and the research data – publicly accessible in digital format with no or minimal restriction" (OECD, 2015:7)

=>

- Primární výstupy veřejně financovaného výzkumu (publikace, výzkumná data)
veřejně dostupné v digitální formátu s žádnými nebo minimálními restrikcemi



Open science jako současný trend vědní politiky

- Poskytovatelé podpory VaV vytvářejí podmínky pro financování navázané na koncepty OS
- Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on open data and the re-use of public sector information
- <https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1561563110433&uri=CELEX:32019L1024>

Zákon č. 130/2002 Sb.

- § 12a Přístup k výzkumným datům
- novela 30. 8. 2022 transponuje směrnici EU o otevřených datech a opakovaném použití informací veřejného sektoru.
- Změna platí pro všechny veřejné soutěže, mezinárodní výzvy a veřejné zakázky ve výzkumu, vývoji a inovacích vyhlášené po 1. 9. 2022.
- Příjemci mají nově povinnost zpracovávat Plán správy dat, ve kterém bude uveden způsob správy výzkumných dat a informace o dostupnosti a způsobu šíření výsledků výzkumu a výzkumných dat.

Open Science není jen Open data, ale i

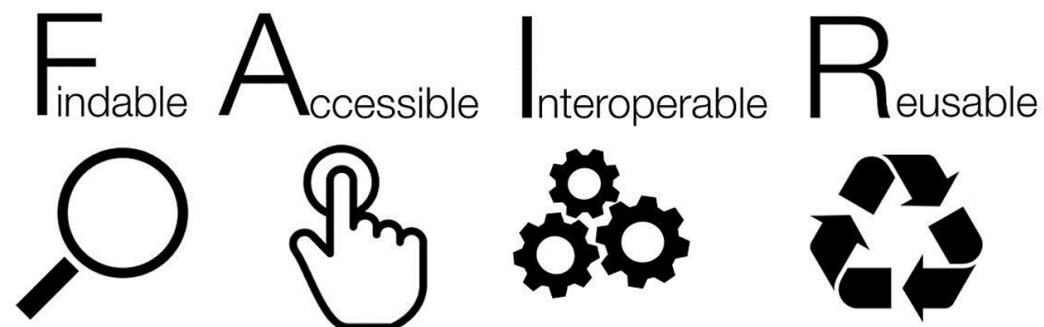
- Open Access, Open Source, Open Licensing, Open educational resources, Open notebooks

Cíle Open Science:

- 1) umožnit každému přístup k vědeckým výstupům, ať už jednotlivcům z vědecké komunity, ale i laikům
- 2) podpořit a zesílit vědeckou kooperaci, výměnu know-how

FAIR Data

- Mantra současného datového managementu
- Není to zcela to samé co Open Access
- FAIR data: „**As open as possible, as closed as necessary**”



FAIROvě s daty a metadaty:

Findability (Vyhledatelnost)

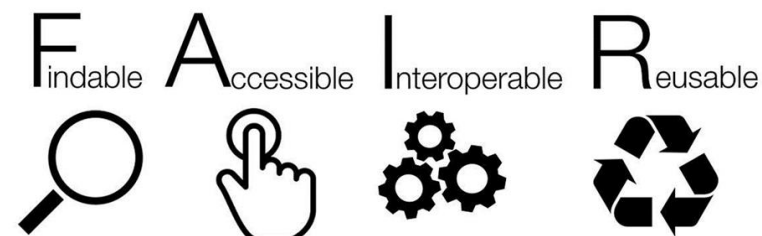
- Mají PID (Persistent identifier; DOI)
- Plně dostačující metadata, která jsou zveřejněna

Accessibility (Přístupnost)

- PID (DOI) lze snadno najít
- Metadata jsou dlouhodobě/navždy dostupná

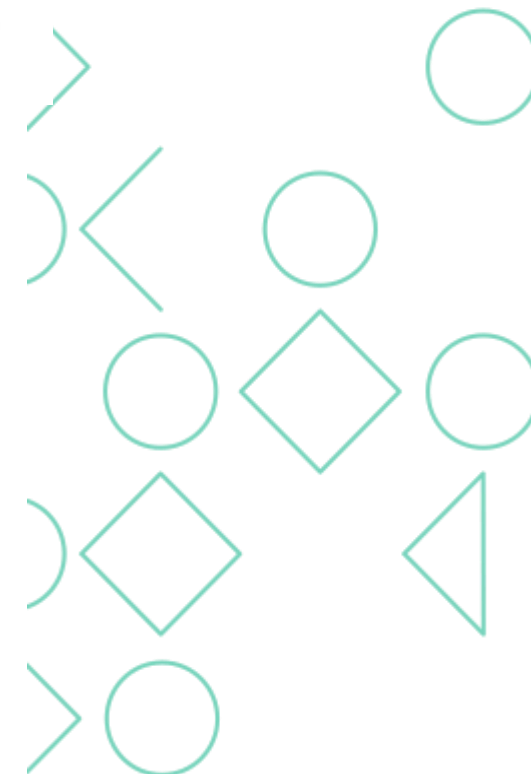
Interoperability

- Formální, komunitě srozumitelný jazyk
- Použití slovníků (Thesaurus CESSDA, ELSST)
- Informace o vztazích k dalším datům



Reuse (Opakované použití)

- Plně dostačující metadata
- Licence
- Celkově následování standardů ve vědní oblasti



Management dat

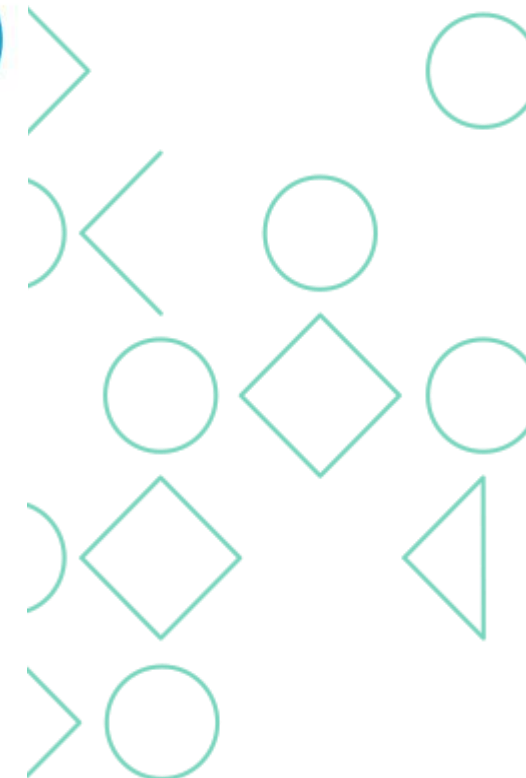
Research data lifecycle =>

- Datový management se pohybuje po tomto kruhu
- DMEG - Data Management Expert Guide
- <https://dmeg.essda.eu/Data-Management-Expert-Guide>



Při přípravě výzkumu je potřeba myslet dopředu

- **Proč?** Některé otázky, které musím řešit už na začátku výzkumu, abych nebyla nemile překvapena
- Jaké metody a nástroje výzkumu použiju?
Budu se inspirovat v již existujících datech?
- Etické náležitosti - ochrana respondentů, budu mít osobní data? Jak s nimi naložím?
- Back-up - jak ochráním data proti ztrátě? V průběhu analýzy a při archivaci?
- Licence - musím mít vyřešeno, kdo a jak bude moci používat moje data
- Archivace a publikování - neusiluji o ně jen pro blaho vědecké komunity, ale i z důvodu nároků poskytovatele podpory



Variabilita DMP

- Oborová diverzita
- Neexistuje jeden daný předpis, jak má DMP vypadat
- Jde o to, zahrnout všechny informace, které jsou relevantní
- V kontextu sociálních věd velmi dobře funguje **DMP DMEG CESSDA**
- Kapitola Plan: <https://dmeg.cessda.eu/Data-Management-Expert-Guide/1.-Plan/Adapt-your-DMP-Part-1>
- **DMP pro kvalitativní data**
- <https://dmeg.cessda.eu/content/download/3536/33233/file/DMPQuestionsQualitativeData.pdf>
- **DMP pro kvantitativní data**
- <https://dmeg.cessda.eu/content/download/3537/33238/file/DMPQuestionsQuantitativeData.pdf>

Další templates:

Horizon Europe

<https://enspire.science/wp-content/uploads/2021/09/Horizon-Europe-Data-Management-Plan-Template.pdf>

Překlad CZ

https://repozitar.techlib.cz/record/1601/files/Horizon_Europe_DMP_template_CZ_EN_version2_0_PDF.pdf

Online nástroje

vyplnění online, odpovídání na otázky

DS Wizard

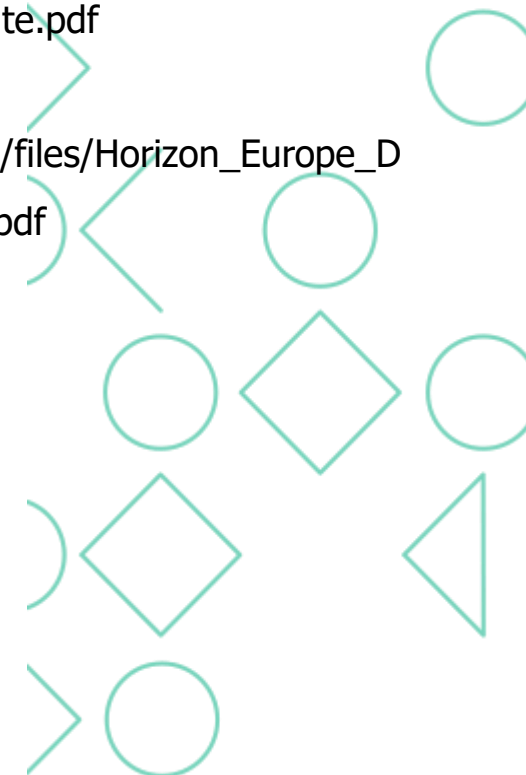
<https://ds-wizard.org/>

DMP online

<https://dmponline.dcc.ac.uk/>

Argos (Open AIRE)

<https://argos.openaire.eu/splash/>



Hand out: CESSDA DMEG DMP Questions

Overview

Title of the project/study

Date of this plan

Description of the project

- What is the nature of the project?
- What is the research question?
- What is the project time line?

Origin of Data

- What kind of data will be used during the project?
- If you are reusing existing data: What is the scope, volume and format? How are different data sources integrated?
- If you are collecting new data can you clarify why this is necessary?

Principal researchers

- Who are the main researchers involved?
- What are their contact details?

Collaborating researchers (if applicable)

- What are their contact details and their roles in the project?

Funder (if applicable)

- If funding is granted, what is the reference number of the funding granted?
- What is the project's title in the funding contract?

Data producer

- Which organisation has the administrative responsibility for the data?

Project data contact

- Who can be contacted about the project during and after it has finished?

Data owner(s)

- Which organisation(s) own(s) the data?
- If several organisations are involved, which organisation owns what data?

Roles

- Who is responsible for updating the DMP and making sure that it's followed?
- Do project participants have any specific roles?
- What is the project time line?

Costs and Resources

- Are there costs you need to consider to buy specific software or hardware?
- Are there costs you need to consider for storage and backup?
- Are potential expenses and resources for (preparing the data for) archiving covered?
- What resources will be dedicated to data management ensuring that data will be FAIR?

Organising and documenting your data

Data collection

- How will the data be collected?
- Is specific software or hardware or staff required?
- Who will be responsible for the data collection?
- During which period will the data be collected?
- Where will the data be collected?

Data organisation

- How will you organise your data?
- Will the data be organised in simple files or more complex databases?
- How will the data quality during the project be ensured?
- If data consists of many different file types (e.g. videos, text, photos), is it possible to structure the data in a logical way?

Data type and size

- What type(s) of data will be collected?
- What is the scope, quantity and format of the material?
- After the project: What is the total amount of data collected (in MB/GB)?

File format

- In what format will your data be?
- Does the format change from the original to the processed/final data?
- Will your (final) data be available in an open format?

Folder structure and names

- How will you structure and name your folders?

File structure and names

- How will you structure and name your files?

Documentation

- What documentation will be created during the different phases of the project?
- How will the documentation be structured?

Metadata

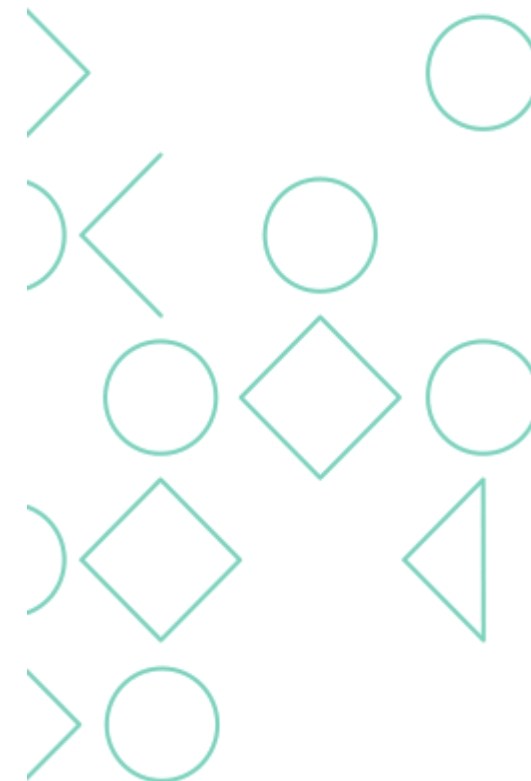
- What metadata will be provided with the collected/ generated/ reused data?
- How will metadata for each object be created?
- Is there any program that can be used to document the data?
- Can metadata be added directly into the files or will the metadata be produced in another program or document?

Metadata standard (if applicable)

- What metadata standard(s) will you use?

Hand out: CESSDA DMEG DMP Questions

- Rozdělen do sedmi hlavních podkategorií - opisuje Data Lifecycle
 - https://dmeg.cessda.eu/content/download/4302/48656/file/TTT_DO_DMPExpertGuide_v1.3.pdf
1. Overview - **Obecné informace**
 2. Organising and documenting your data - **Organizace a dokumentace dat**
 3. Processing your data - **Zpracování dat**
 4. Storing your data and metadata - **Uložení dat**
 5. Protecting your data - **Ochrana dat**
 6. Archiving and publishing your data - **Archivace a publikace dat**
 7. Discovering data - **Vyhledávání dat pro sekundární analýzu**
- Podoba DMP je přizpůsobitelná, není třeba uvádět všechny podrobnosti např. o sekundárních datech (organizace, zpracování), klást důraz na data, která budou vytvořena v rámci projektu



1. Overview - Obecné informace

- Důležité a nutné uvádět vždy
- Název projektu, popis projektu, hlavní výzkumníci, členové výzkumného týmu projektu, poskytovatel podpory, vlastník dat, producent dat, kontakty, přiřazené role v datovém managementu - kdo zpracovává DMP, kdo další se podílí na DM?, náklady na management dat - software, storage + back-up, z jakých zdrojů bude financováno

2. Organising and documenting data - Organizace a dokumentace dat

- Popis sběru dat, organizace dat - rozmyslet jaké typy dat mám a jak je budu kombinovat, pokud mám komplexní data (vzájemně propojené soubory), musím to dopředu promyslet, typ dat, použité formáty, struktura složek, ve kterých jsou data uložena a jejich názvy, struktura a názvy souborů, dokumentace, použitý metadata standard

3. Processing data - Zpracování dat

- Verze, strategie pojmenovávání verzí, velmi důležité!
- Masterfile - kdo je za něj zodpovědný?
- Interoperabilita - přístupnost softwaru, použití slovníků (např. ELSST)
- Kódování - informace o standardizovaných kódech v datech (např. ISCO)
- Informace o kvalitě dat, patří však spíše do metadat - dokumentace dat

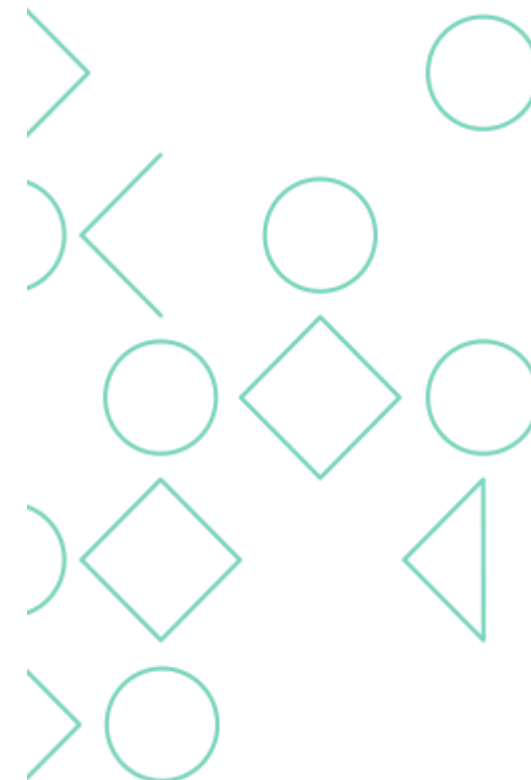


4. Storing data and metadata - Uložení dat

- Kde budou (meta)data uložena v průběhu projektu? Kde bude masterfile? Kde po skončení projektu?
- Jak bude zajištěno zálohování? budou zálohována další data kromě masterfilu? např. editované verze s nově vytvořeným kódováním?
- Jak jsou uložená data chráněna před náhlou ztrátou v důsledku vnějších okolností (povodeň, požár)
- Ochrana dat - jak jsou chráněna osobní data proti vniknutí nepovolaných osob?
- Kdo bude mít přístup k osobním datům?

5. Protecting data – Ochrana dat

- Ethical review - je potřeba? musí výzkum schválit nějaká etická komise (instituce atp.)?
- Informovaný souhlas - bude potřeba? Znění. Jak budou archivovány vyplněné formuláře?
- Osobní data - budou sbírána? Pokud ano, jak budou chráněna a kdo k nim bude mít přístup?
- Copyright - rozhodnutí o podmínkách užití dat
- Dohody/smlouvy s jinými subjekty intervenujícími v projektu (např. agentura pro sběr dat)
- Jakékoliv poznámky k užití dat, restrikcím, podmínkám užití atp.

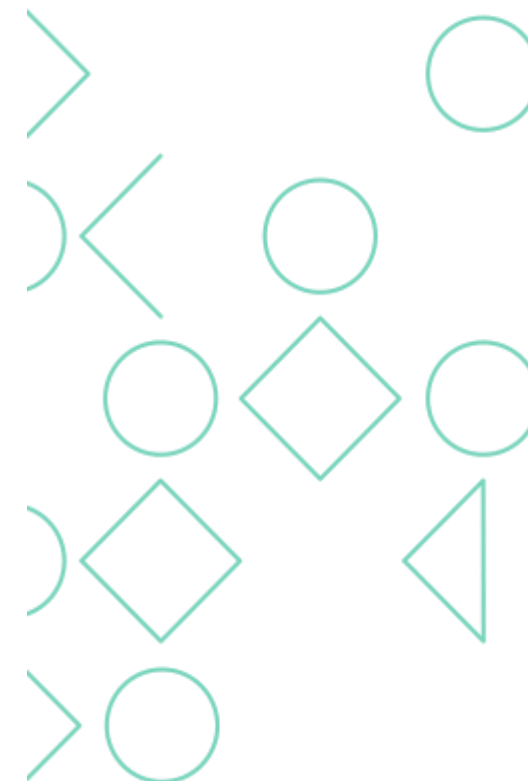


6. Archiving and publishing your data - Archivace a publikace dat

- Kde budou data archivována (archiv, repozitář, instituce), má archiv osvědčení „trusted repository“? Bude datům přidělen PID (DOI)?
- V jakém formátu budou data archivována
- Publikace dat - Přístup k datům
- Celá/část dat publikována, podmínky přístupu, licence užití, citace dat, bude dočasné embargo na re-use dat?

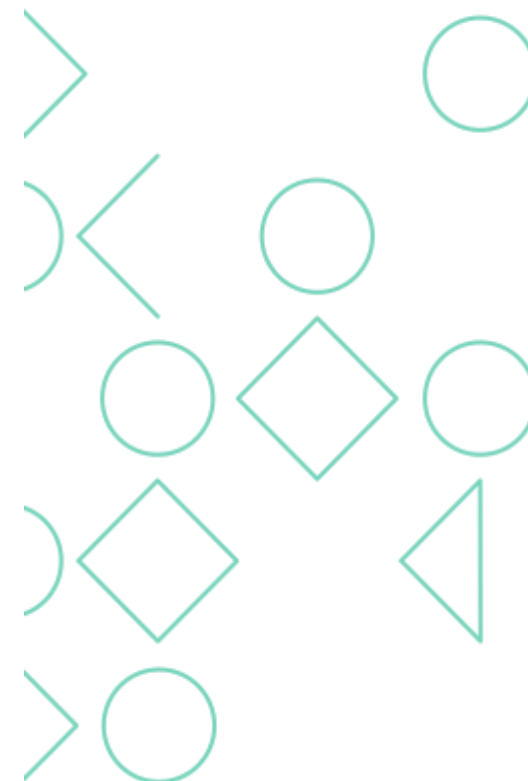
7. Discovering data - Vyhledávání dat pro sekundární analýzu

- Stručně o datech získaných z repozitářů/jiných zdrojů pro sekundární analýzu dat



Důležité odkazy

- Go Fair <https://www.go-fair.org/fair-principles/>
- CESSDA Data Management Expert Guide <https://dmeg.cessda.eu/>
- Foster Open Science <https://www.fosteropenscience.eu/>



Krátce o Data discovery pro sekundární analýzu

Kroky v Data Discovery:

1. Rozmyšlení, jaká data potřebujete
2. Nalezení vhodného zdroje dat
3. Vyhledávání dat v datovém zdroji (archivu)
4. Výběr vhodných dat
5. Zhodnocení kvality nalezených dat

DMEG Data Discovery Kapitola

obsahuje detailní seznam datových repozitářů/archivů a dalších zdrojů dat a ukazuje, kam se obracet

1. Archivy CESSDA

2. Registry datových repozitářů - Re3data.org, OpenDOAR etc.
3. Vyhledávač dat - Google dataset search
4. Agregátory metadat - DataCite, DataSearch

5. CESSDA Data catalogue - data ze všech CESSDA archivů

6. Datové časopisy - Research Data Journal for the Humanities and Social Sciences, Scientific data



DMEG Data Discovery Kapitola

Seznam důležitých datových zdrojů:

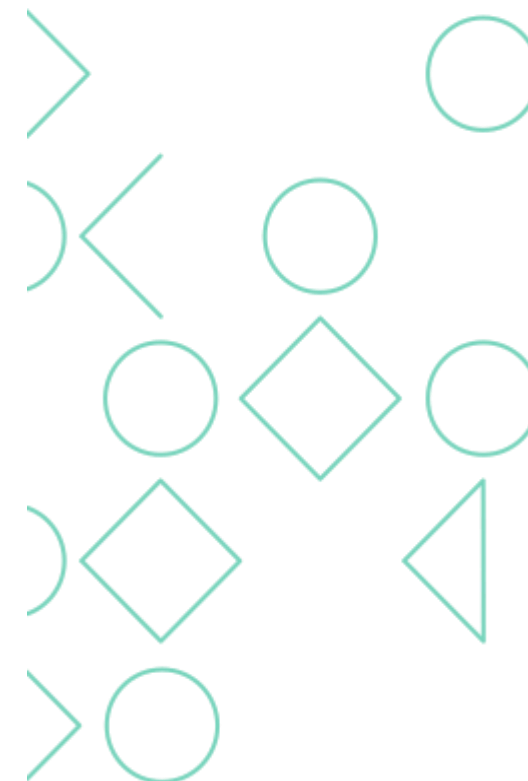
- CESSDA archivy
- Evropské archivy mimo CESSDA
- Vybrané mimo Evropské datové archivy
- Další významné datové repozitáře (Eurostat, OECD stat data, World Bank Open data atd.)

Navíc:

- Datové zdroje k problematice stárnutí
- Datové zdroje k mezinárodnímu komparativnímu výzkumu
- CESSDA Data training, informační a studijní materiály

ELSST - European Language Social Science Thesaurus

- mezinárodně sladěný seznam pojmů, které slouží jako klíčová slova při vyhledávání dat v archivech
- kooperace evropských datových archivů, používán v archivech CESSDA
- <https://elsst.cessda.eu/>



Děkuji za pozornost

johana.chylikova@soc.cas.cz

<https://archiv.soc.cas.cz/cz/>

