

tl;dr: this AI sums up research p. X




https://www.nature.com/articles/d41586-020-03277-2

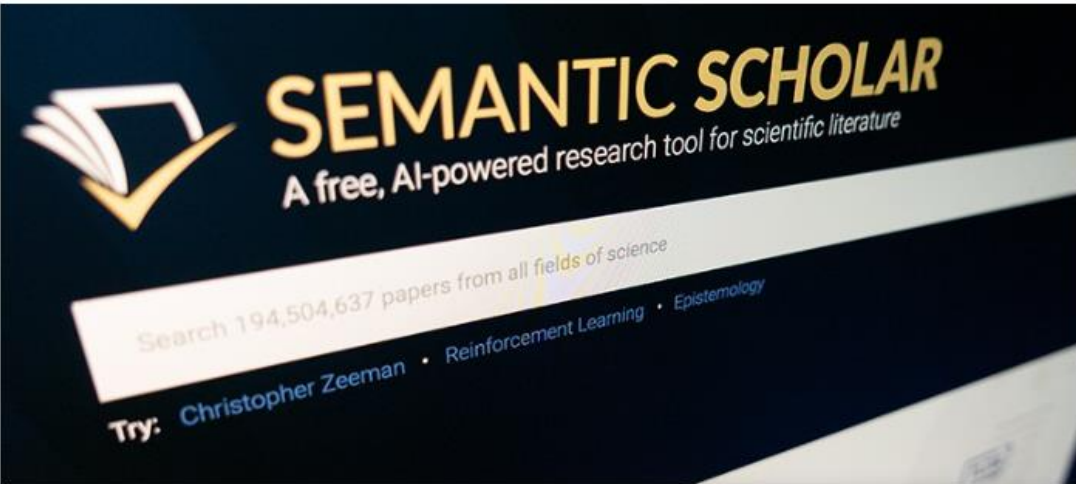
NEWS | 23 November 2020

# tl;dr: this AI sums up research papers in a sentence


Search engine's tool for summarizing studies promises easier skim-reading.


[Jeffrey M. Perkel](#) & [Richard Van Noorden](#)


  





**Related Articles**

[How AI technology can tame the scientific literature](#) 

[AI science search engines expand their reach](#) 

[Coronavirus in context: Scite.ai tracks positive and negative citations for COVID-19 literature](#) 

[Science search engine links papers to grants and patents](#) 

[Artificial-intelligence tools aim to tame the coronavirus literature](#) 

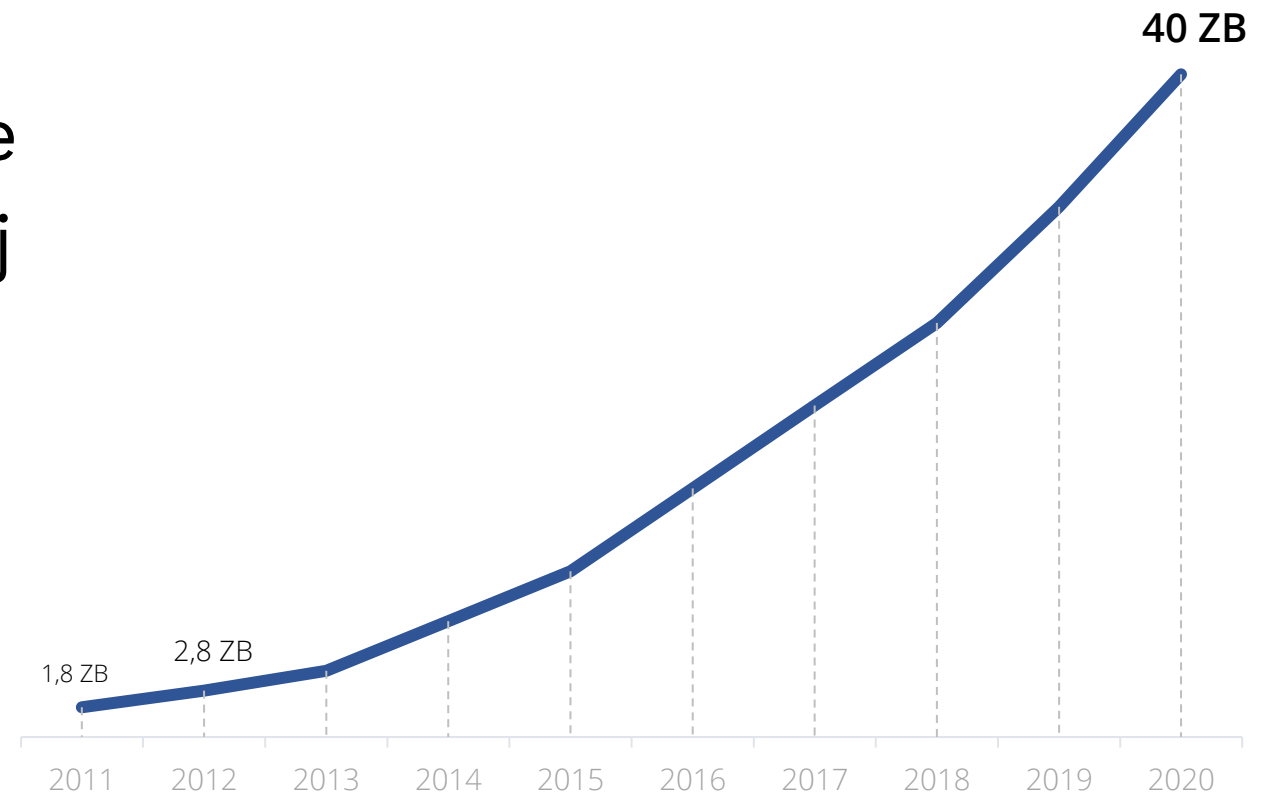
# **Vyhledávání informací**

Hledání dat a v datech

12. 11. 2021

# Data jsou všude... to už víme!

- všechno se měří a ukládá
- mnoho dat je k ničemu
- mnoho se jich zveřejňuje
- některá jsou cenný zdroj



# Data v různých podobách

V čem všem můžeme vyhledávat?

- expertní systémy
  - faktografické a statistické databáze
  - datové repozitáře
  - *+ specifické cesty*
- 
- *data? dataset?*

# 1. Expertní systémy

- napodobují fungování odborníka
- báze znalostí / báze dat / řídicí mechanismus
- pravidla IF-THAN
- dotazování – jako skutečný expert
- postupné hledání výstupu
- podpora rozhodování


## 2. Faktografické databáze

- údajovou základnu tvoří faktografické informace
- *specializované databáze (chemie, fyzika,...)*
- *statistické databáze*
- *encyklopedické databáze*
  
- identifikace zdroje
- hodnocení kvality


Encyclopedia of Life x +

https://eol.org

eol en

  
Encyclopedia of Life

Global access to knowledge about life on Earth



Convergent Lady Beetle   Cliff Chipmunk   ostrich fern   Greater Blue-ringed Octopus   *Nostoc linckia*   *Eucyclops speratus* (Lilljeborg 1901)   Christmas tree worm

Join the discussion! feedback and bio-chatter

Biodiversity Cards View, print or make cards

Data Search nocturnal arboreal frugivores

Trait based models in the VERA modeling tool

Explore foodwebs who's eating whom?

[What is EOL?](#) [What's New?](#) [The EOL Forum](#) [Education](#) [Citing EOL](#) [Terms of Use](#) [Data Services](#) [Food Webs](#) [Language Support](#) [Donate](#) [Contact Us](#)

# Vyhledávačka



Pokuste se co nejrychlejší cestou najít seznam všech druhů ptáků, kteří se vyskytují na území Jihomoravského kraje.



Avibase - The World Bird Database

Bird checklists - taxonomy - distribution - maps - links

Avibase myAvibase Checklists Search Contribute English Sign Up Sign In

Welcome Guest

## Welcome to Avibase

Avibase is an extensive database information system about all birds of the world, containing over 53 million records about 10,000 species and 22,000 subspecies of birds, including distribution information for 20,000 regions, taxonomy, synonyms in several languages and more. This site is managed by Denis Lepage and hosted by Birds Canada, the Canadian copartner of BirdLife International. Avibase has been a work in progress since 1992 and I am now pleased to offer it as a service to the bird-watching and scientific community.

© Denis Lepage 2021 - Number of records currently in Avibase: **53,973,606** - Last update: **2021-10-30**

**Bird of the day:** *Incaezpiza laeta* (Buff-bridled Inca-Finch)

View archives




photo @uropsalis  
Report: [Progress bar] (1 votes)  
Photo powered by flickr.com.  
Listen to recording

BIRDS CANADA OISEAUX CANADA

Canadian co-partner of un partenaire canadien de

BirdLife INTERNATIONAL

**Bird checklist of the day:** Orinoquia (Region), Colombia

eBird - Discover a new world of

The Cornell Lab of Ornithology

eBird Submit Explore My eBird Science About News Help

ZEISS Seeing beyond

Explore - eBird

https://ebird.org/explore

eBird Submit Explore My eBird Science About News Help

Donate Create account Sign in Language

## Explore Regions

Recent sightings, checklists, birding activity, best hotspots, and top birders for a county, state, province, country, territory, or dependency

Enter a region


National Wildlife Refuges Major Regions

### Explore Hotspots


Discover the best places for birding nearby or around the world.

### Bar Charts

Find out what birds to expect throughout the year in a region or location




*Tyrannus melancholicus* © James MacKenzie Macaulay Library eBird



FOODB [Browse](#) [Search](#) [Downloads](#) [Reports](#) [Examples](#) [About](#) [Contact Us](#)

**TMIC** The **Metabolomics** Innovation Centre | Quantitative metabolomics services for biomarker discovery and validation.



# FOODB

[Browse Foods >>](#)

[Browse Compounds >>](#)

Welcome to FooDB Version 1.0

WISQARS (Web-based Injury St...

https://www.cdc.gov/injury/wisqars/index.html

**CDC** Centers for Disease Control and Prevention  
CDC 24/7: Saving Lives, Protecting People™

[A-Z Index](#)

Search

[Advanced Search](#)

**Injury Prevention & Control**

Injury Center

[f](#) [t](#) [in](#) [RSS](#)

# WISQARS™

## Injury Data

### WISQARS™ — Web-based Injury Statistics Query and Reporting System

CDC's WISQARS™ is an interactive, online database that provides fatal and nonfatal injury, violent death, and cost of injury data. Researchers, the media, public health professionals, and the public can use WISQARS™ data to learn more about the public health and economic burden associated with unintentional and violence-related injury in the United States.

Unintentional Suicide

Fatal Injury and

Nonfatal Injury

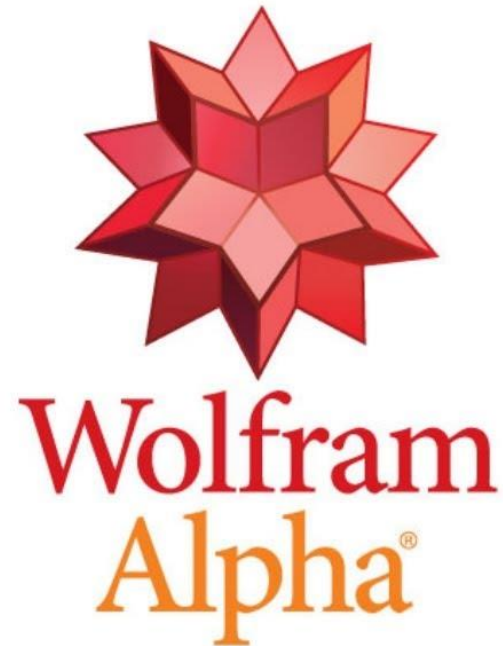
10 Leading Causes of Death by Age Group, United States - 2018


Rank	1-4	5-9	10-14	15-19	20-24	25-29	30-34	35-39	40-44	45-49	50-54	55-59	60-64	65-69	70-74	75-79	80+	Total	
1	Heart Disease	Heart Disease	Heart Disease	Heart Disease	Heart Disease	Heart Disease	Heart Disease	Heart Disease	Heart Disease	Heart Disease	Heart Disease	Heart Disease	Heart Disease	Heart Disease	Heart Disease	Heart Disease	Heart Disease	Heart Disease	Heart Disease
2	Cancer	Cancer	Cancer	Cancer	Cancer	Cancer	Cancer	Cancer	Cancer	Cancer	Cancer	Cancer	Cancer	Cancer	Cancer	Cancer	Cancer	Cancer	Cancer
3	Accidents (unintentional injuries)	Accidents (unintentional injuries)	Accidents (unintentional injuries)	Accidents (unintentional injuries)	Accidents (unintentional injuries)	Accidents (unintentional injuries)	Accidents (unintentional injuries)	Accidents (unintentional injuries)	Accidents (unintentional injuries)	Accidents (unintentional injuries)	Accidents (unintentional injuries)	Accidents (unintentional injuries)	Accidents (unintentional injuries)	Accidents (unintentional injuries)	Accidents (unintentional injuries)	Accidents (unintentional injuries)	Accidents (unintentional injuries)	Accidents (unintentional injuries)	
4	Stroke	Stroke	Stroke	Stroke	Stroke	Stroke	Stroke	Stroke	Stroke	Stroke	Stroke	Stroke	Stroke	Stroke	Stroke	Stroke	Stroke	Stroke	Stroke
5	Diabetes	Diabetes	Diabetes	Diabetes	Diabetes	Diabetes	Diabetes	Diabetes	Diabetes	Diabetes	Diabetes	Diabetes	Diabetes	Diabetes	Diabetes	Diabetes	Diabetes	Diabetes	Diabetes
6	Chronic lower respiratory diseases	Chronic lower respiratory diseases	Chronic lower respiratory diseases	Chronic lower respiratory diseases	Chronic lower respiratory diseases	Chronic lower respiratory diseases	Chronic lower respiratory diseases	Chronic lower respiratory diseases	Chronic lower respiratory diseases	Chronic lower respiratory diseases	Chronic lower respiratory diseases	Chronic lower respiratory diseases	Chronic lower respiratory diseases	Chronic lower respiratory diseases	Chronic lower respiratory diseases	Chronic lower respiratory diseases	Chronic lower respiratory diseases	Chronic lower respiratory diseases	
7	Alzheimer's disease	Alzheimer's disease	Alzheimer's disease	Alzheimer's disease	Alzheimer's disease	Alzheimer's disease	Alzheimer's disease	Alzheimer's disease	Alzheimer's disease	Alzheimer's disease	Alzheimer's disease	Alzheimer's disease	Alzheimer's disease	Alzheimer's disease	Alzheimer's disease	Alzheimer's disease	Alzheimer's disease	Alzheimer's disease	
8	Chronic kidney disease	Chronic kidney disease	Chronic kidney disease	Chronic kidney disease	Chronic kidney disease	Chronic kidney disease	Chronic kidney disease	Chronic kidney disease	Chronic kidney disease	Chronic kidney disease	Chronic kidney disease	Chronic kidney disease	Chronic kidney disease	Chronic kidney disease	Chronic kidney disease	Chronic kidney disease	Chronic kidney disease	Chronic kidney disease	
9	Septicemia	Septicemia	Septicemia	Septicemia	Septicemia	Septicemia	Septicemia	Septicemia	Septicemia	Septicemia	Septicemia	Septicemia	Septicemia	Septicemia	Septicemia	Septicemia	Septicemia	Septicemia	
10	Intentional self-harm (suicide)	Intentional self-harm (suicide)	Intentional self-harm (suicide)	Intentional self-harm (suicide)	Intentional self-harm (suicide)	Intentional self-harm (suicide)	Intentional self-harm (suicide)	Intentional self-harm (suicide)	Intentional self-harm (suicide)	Intentional self-harm (suicide)	Intentional self-harm (suicide)	Intentional self-harm (suicide)	Intentional self-harm (suicide)	Intentional self-harm (suicide)	Intentional self-harm (suicide)	Intentional self-harm (suicide)	Intentional self-harm (suicide)	Intentional self-harm (suicide)	

Leading Causes of Death and

# Wolfram Alpha

<https://www.wolframalpha.com/>




 **WolframAlpha** computational... knowledge engine

International Space Station May 18 2009 6:00pm

Input interpretation:

International Space Station (spacecraft)	6:00:00 pm CDT   Monday, May 18, 2009
---	--

Position at 6:00 pm: [Show DMS](#) | [Show 3D](#) | [Orthographic projection](#)



22.51° North 94.35° West (ocean)

Orbital information at 6:00 pm: [Show metric units](#) | [More](#)

# Dotazovací jazyk Wolfram

- operátory *vs*, *>*, *<*, *=*, *in*,...
- thorin vs. frodo
- zpracování čísel
- 85 kg 192 cm 12 beers in 5 hours
- zpracování přirozeného jazyka
- libraries with number of books *>* 15 million
- programovací jazyk Wolfram
- `Range[x]`,...

# Statistické databáze

- obrovské množství databází mezinárodních organizací
- databáze státních orgánů, úřadů
- většinou vyhledávací rozhraní
- mnohdy možnosti vizualizace
- často omezené možnosti práce s *raw* daty

<https://data.worldbank.org/>  
<https://www.who.int/data/collections> <-  
<https://data.unicef.org/>  
<https://stats.oecd.org/>  
<http://data.europa.eu/>

Český statistický úřad | ČSÚ

https://www.czso.cz

ČESKÝ STATISTICKÝ ÚŘAD

Kontakty Nejžádanější Časté dotazy Nápoředa


Statistiky Vydáváme **Databáze, registry** Klasifikace, číselníky Výkazy, sběr dat O ČSÚ

- > [Veřejná databáze](#)
- > [Databáze národních účtů](#)
- > [Databáze demografických údajů za obce ČR](#)
- > [Databáze demografických údajů za vybraná města ČR](#)
- > [Registr ekonomických subjektů, IČO](#)
- > [Databáze pohybu zboží přes hranice](#)
- > [Registr osob](#)
- > [Registr sč. obvodů a budov](#)
- > [Ubytovací zařízení ČR](#)
- > [Databáze Eurostatu v češtině](#)

Nejnovější údaje

<a href="#">Obyvatelstvo</a>	10 702 942 ↑
<a href="#">Průměrná hrubá mzda</a>	38 275 Kč ↑
<a href="#">Míra inflace</a>	3,0 % ↑
<a href="#">Hrubý domácí produkt</a>	2,8 % ↑
<a href="#">slová produkce</a>	-4,0 % ↓

Vyberte si kraj



Více informací

09.11. [Cestovní ruch - 3. čtvrtletí 2021](#)

08.11. [Zahraniční obchod se zbožím - září 2021](#)

08.11. [Průmysl - září 2021](#)

08.11. [Stavebnictví - září 2021](#)

05.11. [Epidemie covid-19 zasáhla maloobchod ve třech vlnách](#)

https://vdb.czso.cz/vdbvo2/

# 3. Repozitáře dat

- úložiště dat a **metadat**
- data většinou v podobě datových souborů
- možnost stáhnout a dále analyzovat, zpracovávat
- velké množství repositářů *otevřených dat*



# OPEN DATA

Otevřená data jsou informace a data bezplatně a volně dostupná na internetu ve strukturované a strojově čitelné podobě a jsou zpřístupněna způsobem, který jejich využití neklade zbytečné technické či jiné překážky.

Portál otevřených dat České rep. X +

← → ↻ 🏠 🔒 https://data.gov.cz

OTEVŘENÁ DATA

Vzdělávání Články Datová kvalita Národní katalog otevřených dat Další informace 🇨🇪

# Portál otevřených dat

Vše, co potřebujete vědět o otevřených datech v České republice

Národní katalog otevřených dat

2 658 klíčových slov

🔍 Název datové sady Hledat

## Novinky ze světa otevřených dat

<p>Právě probíhá výroční konference Otevřená data 2021</p> <p>5. 11. 2021</p>	<p>Nová zákonná povinnost – „publikace úředních desek jako otevřená data“</p> <p>4. 11. 2021</p>	<p>Registrujte se na konferenci Otevřená data 2021</p> <p>18. 10. 2021</p> <p>Konference Otevřená data 2021 se letos uskuteční 5. listopadu a bude probíhat</p>
---	--	---

https://data.gov.cz/datové-sady 5 hodin dnes probíhá v Národní Dne 16. 11. 2021 proběhne [speciální](#)


data.brno

https://data.brno.cz

Sign In

DATA APLIKACE ANALÝZY ČLÁNKY ZPRÁVA O STAVU MĚSTA O WEBU

# Vítejte na datovém portálu města Brna!



VSTUP DO KATALOGU

Vyhledej...

Př

Klikněte pro dlouhodobě sledované statistiky...

- EKONOMIKA A TRH PRÁCE
- ŽIVOTNÍ PROSTŘEDÍ
- DOPRAVA
- LIDÉ A BYDLENÍ
- VZDĚLÁNÍ
- ZDRAVÍ
- BEZPEČNOST
- MĚSTO

# + **Specifické cesty**

- někde to jen tak leží...
- metavyhledávače
- webscraping

EVIDENCE KNIHOVEN, adresář

https://www.mkcr.cz/evidence-knihoven-adresar-knihoven-vidovanych-m

MINISTERSTVO KULTURY

MINISTERSTVO STRUKTURA DOTACE KKO NOVINKY A MÉDIA KONTAKTY

Úvodní stránka / Struktura / Literatura a knihovny  
/ EVIDENCE KNIHOVEN, adresář knihoven evidovaných Ministerstvem kultury a související informace

## EVIDENCE KNIHOVEN, adresář knihoven evidovaných Ministerstvem kultury a související informace

**Kulturní dědictví**

**Profesionální umění**

**Literatura a knihovny**

**Média a audiovize**

**Autorské právo**

**Zahraniční vztahy**

**Evropské záležitosti**

**Databáze knihoven, informace o evidenci knihoven, vzory formulářů**

- [Databáze knihoven](#) (aktualizace 8. 4. 2021)

**Soubory ke stažení**

[Formulář žádosti o změnu údajů v evidenci knihoven - fyzické osoby](#) (.docx, 16.7 kB)

[Formulář žádosti o změnu údajů v evidenci knihoven - právnické osoby](#) (.docx, 19.8 kB)

[Formulář žádosti o evidenci knihovny - fyzické osoby](#) (.docx, 16.3 kB)

[Formulář žádosti o evidenci knihovny - právnické osoby](#) (.docx, 16.5 kB)

https://www.mkcr.cz/evidence-knihoven-adresar-knihoven-vidovanych-ministerstvem-kultury-a-souvisejici-informace-341.html#null

Počty podnikatelů dle občanství x +

https://www.mpo.cz/cz/podnikani/zivnostenske-podnikani/statisticke-uc

Kontakty EN f t in YouTube

MINISTERSTVO PRŮMYSLU A OBCHODU

Hledaný text ...

Rozcestník Podnikání Zahraniční obchod Ochrana spotřebitele Energetika Stavebnictví a suroviny Průmysl e-Komunikace a pošta

Úvodní stránka / Podnikání / Živnostenské podnikání / Statistické údaje o podnikatelích

## Počty podnikatelů dle občanství podnikajících v České republice

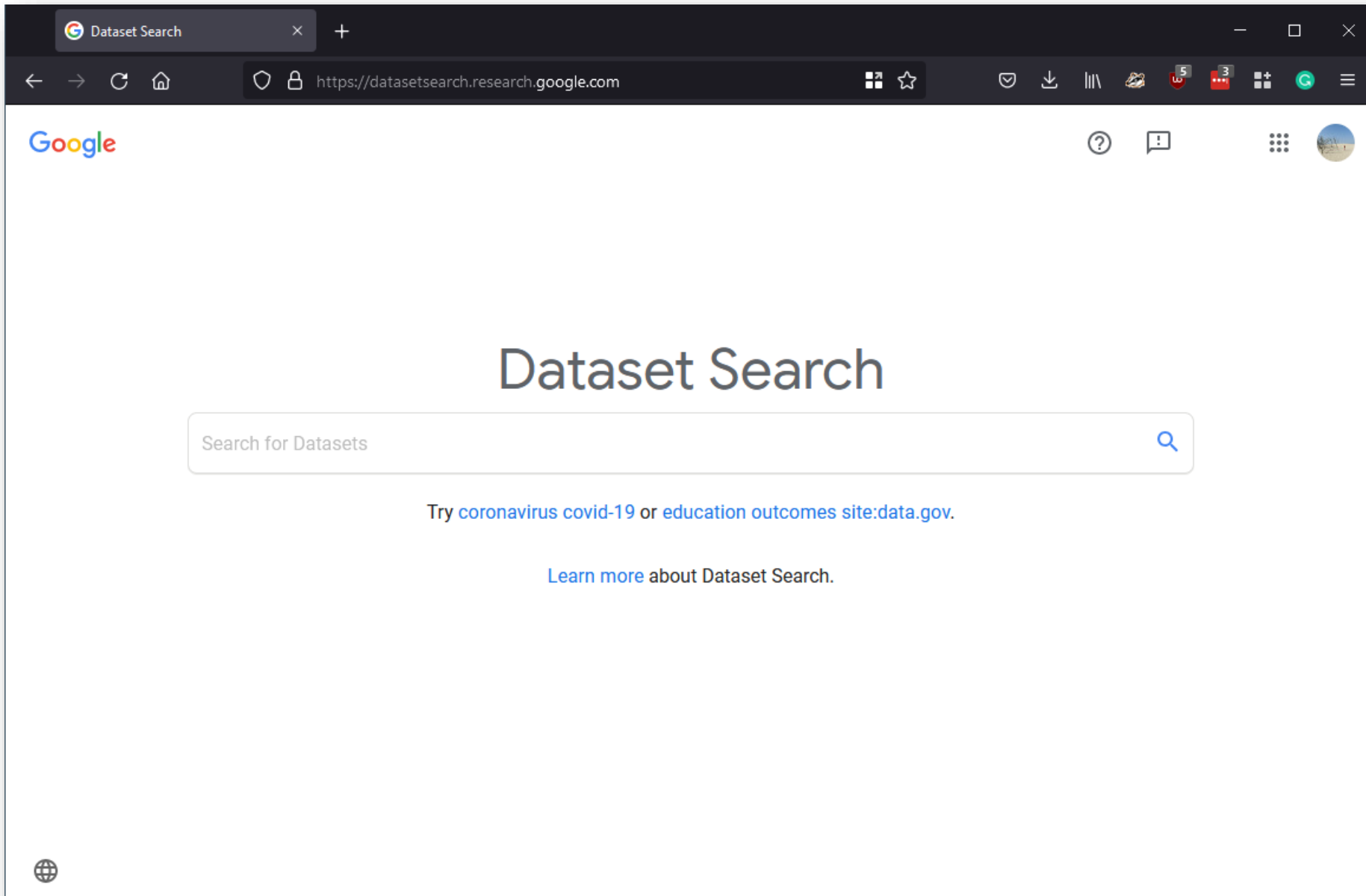
Publikováno: **12.10.2021**  
Autor: **81300**

*Počty podnikatelů dle občanství podnikajících v České republice*

Na základě usnesení vlády České republiky č. 344 ze dne 10. května 2010, bod III/8., zveřejňujeme statistické údaje o působení cizinců podnikajících dle zákona č. 455/1991 Sb., (živnostenský zákon) v České republice, v rozdělení podle státního občanství.

- Národní plán obnovy >
- Živnostenské podnikání ▾
- Právní předpisy
- Živnostenský rejstřík
- Průvodce živnostenským podnikáním
- CRM – Jednotný registrační formulář
- Povinnosti osob dočasně

podnikatelé *filetype:xlsx*





# Vyhledávačka



Dobrý den, potřeboval bych vědět,  
kolik dětí bylo v letošním školním roce  
zapsáno v mateřských školkách v Brně.

# Webscraping

- *web data extraction*, sbírání dat z webu
- vytváření datasetů nad webovými stránkami
- **automatizovaně**
- má to mnoho specifik, podíváme se na to v NaMI



**Strategie hledání relevantních datasetů:** kombinace webového vyhledávače, správných klíčových slov a identifikace relevantních organizací a institucí spojená s prolézáním jejich webových sídel.

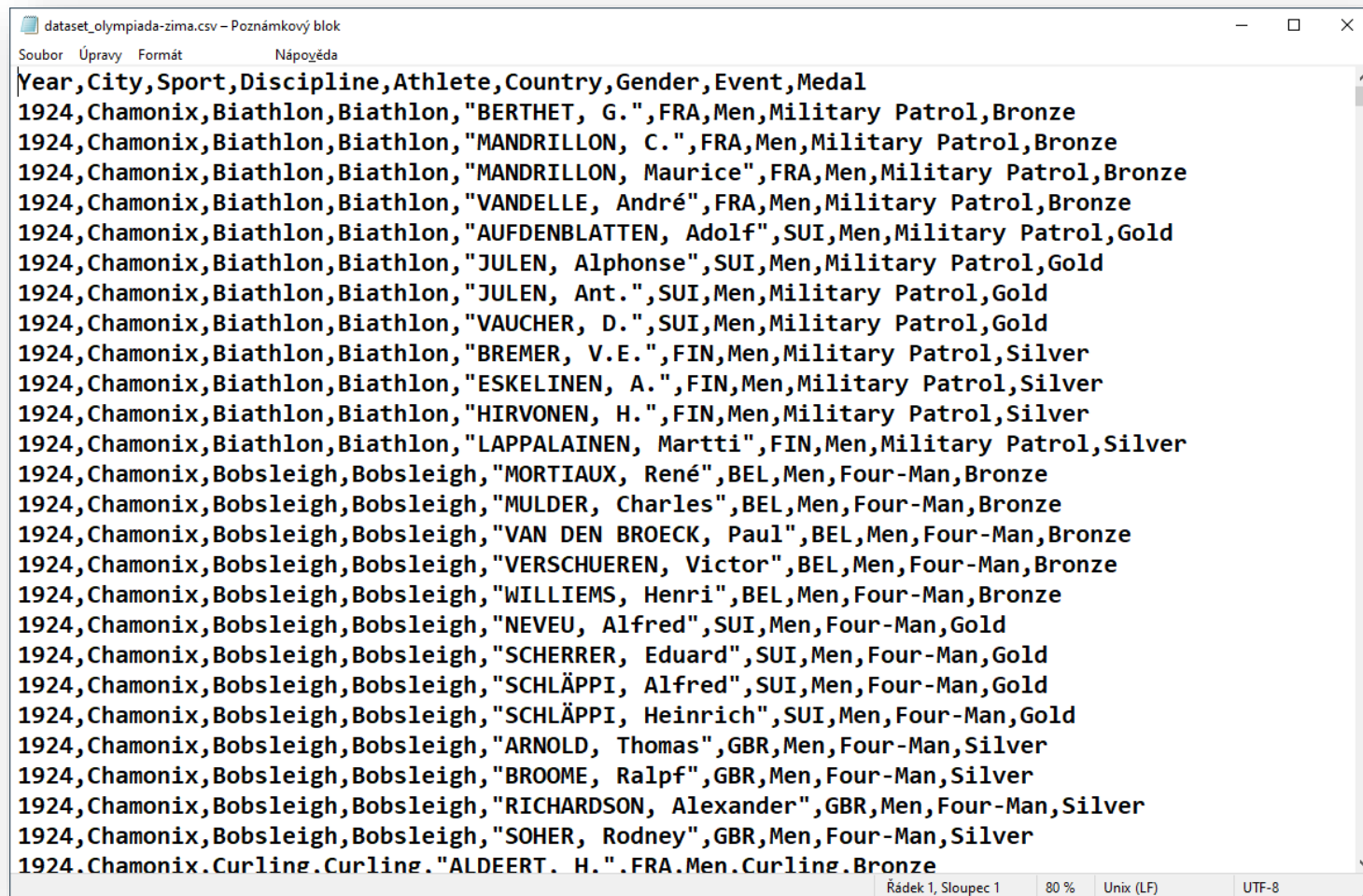
- 1. Identifikujte klíčová slova** k tématu + *data, dataset, statistiky,...*
- 2.** Zkuste prohledat [Národní katalog otevřených dat](#)
- 3. Identifikujte organizace**, které by k řešenému tématu mohly být relevantní (*ministerstva, úřady, oborové organizace a svazy, neziskovky atp.*) a hledejte u nich datové zdroje.
- 4.** Využijte univerzální datové portály nebo datové metavyhledávače

# Typy dat

- způsob uložení dat do souboru
- data.**XLSX**
- data.**CSV**, data.**TAB**
- data.JSON
- data.KML, data.GeoJSON – geografický rozměr

# Typy dat | CSV

- comma  
*separated value*
- oddělovač ale může být kdeco
- čárka, středník, dvojtečka, tabulátor...



```
dataset_olympiada-zima.csv - Poznámkový blok
Soubor Úpravy Formát Nápověda
Year, City, Sport, Discipline, Athlete, Country, Gender, Event, Medal
1924, Chamonix, Biathlon, Biathlon, "BERTHET, G.", FRA, Men, Military Patrol, Bronze
1924, Chamonix, Biathlon, Biathlon, "MANDRILLON, C.", FRA, Men, Military Patrol, Bronze
1924, Chamonix, Biathlon, Biathlon, "MANDRILLON, Maurice", FRA, Men, Military Patrol, Bronze
1924, Chamonix, Biathlon, Biathlon, "VANDELLE, André", FRA, Men, Military Patrol, Bronze
1924, Chamonix, Biathlon, Biathlon, "AUFDENBLATTEN, Adolf", SUI, Men, Military Patrol, Gold
1924, Chamonix, Biathlon, Biathlon, "JULEN, Alphonse", SUI, Men, Military Patrol, Gold
1924, Chamonix, Biathlon, Biathlon, "JULEN, Ant.", SUI, Men, Military Patrol, Gold
1924, Chamonix, Biathlon, Biathlon, "VAUCHER, D.", SUI, Men, Military Patrol, Gold
1924, Chamonix, Biathlon, Biathlon, "BREMER, V.E.", FIN, Men, Military Patrol, Silver
1924, Chamonix, Biathlon, Biathlon, "ESKELINEN, A.", FIN, Men, Military Patrol, Silver
1924, Chamonix, Biathlon, Biathlon, "HIRVONEN, H.", FIN, Men, Military Patrol, Silver
1924, Chamonix, Biathlon, Biathlon, "LAPPALAINEN, Martti", FIN, Men, Military Patrol, Silver
1924, Chamonix, Bobsleigh, Bobsleigh, "MORTIAUX, René", BEL, Men, Four-Man, Bronze
1924, Chamonix, Bobsleigh, Bobsleigh, "MULDER, Charles", BEL, Men, Four-Man, Bronze
1924, Chamonix, Bobsleigh, Bobsleigh, "VAN DEN BROECK, Paul", BEL, Men, Four-Man, Bronze
1924, Chamonix, Bobsleigh, Bobsleigh, "VERSCHUEREN, Victor", BEL, Men, Four-Man, Bronze
1924, Chamonix, Bobsleigh, Bobsleigh, "WILLIEMS, Henri", BEL, Men, Four-Man, Bronze
1924, Chamonix, Bobsleigh, Bobsleigh, "NEVEU, Alfred", SUI, Men, Four-Man, Gold
1924, Chamonix, Bobsleigh, Bobsleigh, "SCHERRER, Eduard", SUI, Men, Four-Man, Gold
1924, Chamonix, Bobsleigh, Bobsleigh, "SCHLÄPPI, Alfred", SUI, Men, Four-Man, Gold
1924, Chamonix, Bobsleigh, Bobsleigh, "SCHLÄPPI, Heinrich", SUI, Men, Four-Man, Gold
1924, Chamonix, Bobsleigh, Bobsleigh, "ARNOLD, Thomas", GBR, Men, Four-Man, Silver
1924, Chamonix, Bobsleigh, Bobsleigh, "BROOME, Ralph", GBR, Men, Four-Man, Silver
1924, Chamonix, Bobsleigh, Bobsleigh, "RICHARDSON, Alexander", GBR, Men, Four-Man, Silver
1924, Chamonix, Bobsleigh, Bobsleigh, "SOHER, Rodney", GBR, Men, Four-Man, Silver
1924, Chamonix, Curling, Curling, "ALDEERT, H.", FRA, Men, Curling, Bronze
Řádek 1, Sloupec 1 80 % Unix (LF) UTF-8
```

dataset\_olympiada-zima.xlsx

dataset\_olympiada-zima.csv

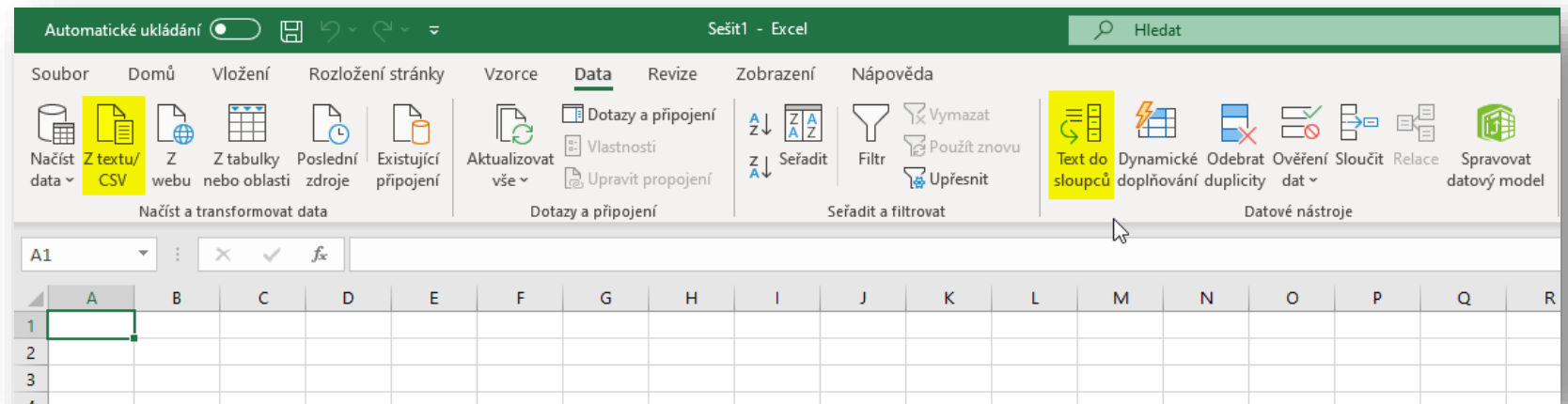
```
dataset_olympiada-zima.csv - Poznámkový blok
Soubor Úpravy Formát Zobrazení Nápověda
Year, City, Sport, Discipline, Athlete, Country, Gender, Event, Medal
1924, Chamonix, Biathlon, Biathlon, "BERTHET, G.", FRA, Men, Mi
1924, Chamonix, Biathlon, Biathlon, "MANDRILLON, C.", FRA, Men,
1924, Chamonix, Biathlon, Biathlon, "MANDRILLON, Maurice", FRA
1924, Chamonix, Biathlon, Biathlon, "VANDELLE, André", FRA, Men
1924, Chamonix, Biathlon, Biathlon, "AUFDENBLATTEN, Adolf", SU
1924, Chamonix, Biathlon, Biathlon, "JULEN, Alphonse", SUI, Men
1924, Chamonix, Biathlon, Biathlon, "JULEN, Ant.", SUI, Men, Mi
1924, Chamonix, Biathlon, Biathlon, "VAUCHER, D.", SUI, Men, Mi
1924, Chamonix, Biathlon, Biathlon, "BREMER, V.E.", FIN, Men, Mi
1924, Chamonix, Biathlon, Biathlon, "ESKELINEN, A.", FIN, Men, M
1924, Chamonix, Biathlon, Biathlon, "HIRVONEN, H.", FIN, Men, M
1924, Chamonix, Biathlon, Biathlon, "LAPPALAINEN, Martti", FIN
1924, Chamonix, Bobsleigh, Bobsleigh, "MORTIAUX, René", BEL, Me
1924, Chamonix, Bobsleigh, Bobsleigh, "MULDER, Charles", BEL, M
1924, Chamonix, Bobsleigh, Bobsleigh, "VAN DEN BROECK, Paul", B
1924, Chamonix, Bobsleigh, Bobsleigh, "VERSCHUEREN, Victor", B
1924, Chamonix, Bobsleigh, Bobsleigh, "WILLIEMS, Henri", BEL, Men, Four-man, bronze
1924, Chamonix, Bobsleigh, Bobsleigh, "NEVEU, Alfred", SUI, Men, Four-Man, Gold
1924, Chamonix, Bobsleigh, Bobsleigh, "SCHEPPER, Edward", SUI, Men, Four-Man, Gold
```

```
dataset_olympiada-zima.xlsx - Poznámkový blok
Soubor Úpravy Formát Zobrazení Nápověda
?9LěÑ"ČsbgŮ|Č!óúUSh9i°bžr:"y_d1RóD>ž|ý|-NsČR"4||ř2DĎGÇ% óZ´4nĚtyÄ7 Ā«ČdÉ,<"T|| ··||
PK|||| | | ! 4JL||Ī|| , || x1/workbook.xml-U]nă6||~/Đ;"B€}R$ĚŎ/b/1IF
$< n&/|| F°#Ā°"RTě4ŘĀě||zŠEdŎ~19öfQ. Ů||||%°tZĚ|3<ú. .||i%Š+nj°tSK×h•ňSÚZCYó|jřšÖHReκr||ēĎ´N?
Ž~ýÍlLĹňóA|| U3Ŏs)ěĐ4>4$%iNyM+ŘYpQ|| Snh6μ $krJeYġeaƒIXAq||Bq||_XJcžġ%-ă||ĐĐ,H Bă~nz
´2=||°$bŮŎFĚĚ|| XÁäs||$ke||Ī+. ČC||nŽ´t-||ü\řG||v ||1°9$đ@r
_ČS€67κBřŽ,||°||-BČr8$1
úĀT||w-,úNVi||Ě}||CŎŎ!|VŠ•||,÷N4gÇÍŎGg||VĐ>Ťt5R×zHŎ2UčZA||"dLŇ1"{}Ŏí+z° ÚzŇ. vměŘžnŽvrI||ZF||κ-ă
„ŮĀCe.n` ;Ě||,1.$||4â•||nYúYÍuRQĪAáŮ||ýte, BaIŔW||I||' #ć' Č
\kE1Ŏtđis||iBÍyúieB_μ||"čň. |ĪrňúnOzám1ü||...°Tām,β||n>÷đc ||EŘ«đR
Tgn9dâš<A^ ůŮġlg||x°te||G±#Đ|| ġ||öbŮđă||6sŔCéĀňČŘú ^||7L9ieIÍμĀę||úfě,-úd...-Ě^Ď||±ġ||+z~7ô
Řádek 1, Sloupec 1 40 % Macintosh (CR) ANSI
```

```
Řádek 1, Sloupec 1 40 % Unix (LF) UTF-8
```

# Základní dovednosti

- otevřít CSV, XLSX a další běžné formáty
- převést data do sloupců v případě CSV
- provést základní operace
- filtrovat si data atp.



# Společná vyhledávačka



Jeden kamarád z Ostravy mi tvrdil, že nejvíc přestupků na území Ostravy páchají Brňáci. To je asi hloupost ne?

**Nedá se to nějak zjistit?**



# Vyhledávačka na doma



Na [Evropském portále otevřených dat](#) najdete data o sankcích EU proti osobám (*sanctions AND person*). Vyberte si jeden z formátů dat, který znáte a bude pro vás přístupný. (CSV, HTML, PDF?) Zjistěte jméno alespoň jedné osoby z Nikaragui, na kterou jsou ze strany EU uvaleny finanční sankce.

# Vyhledávačka na doma



Stáhněte si [tento dataset](#) s přehledem udělených medailí na zimních OH.

Otevřete ho v Excelu a převed'te ho z CSV na funkční tabulku, kde budete moci filtrovat. Pomocí filtru zjistěte, kolik bylo celkem uděleno **zlatých medailí**.

# Data ve vědě

- repozitáře vědeckých dat
- specifické souborové formáty
- weby autorů
- datové časopisy

TOTO TÉMA  
DOKONČENO  
V RÁMCI SETKÁNÍ  
ČÍSLO 9

# Rešerše dat

- Jaká data by měl dataset obsahovat?
- Jaké proměnné, aby byl užitečný?
- Jaký rozsah dat, záběr?
- Jak data vznikla a jakými úpravami prošla?
- V jakých formátech? Pro jaký SW?
- V jaké licenci by měl být dataset?
- Pro komerční nebo nekomerční využití?



# Rešerše dat | Datové repozitáře

- Oborové  
(např. adresář [re3data](#))
- Všeobecné  
(např. [Zenodo](#) nebo [Dryad](#))
- Institucionální  
(např. [Harvard Dataverse](#))



# Vyhledávačka



Využijte adresář datových repozitářů [re3data](#) a pokuste se identifikovat **český** datový repozitář pro oblast **sociálních věd**.

# Rešerše dat | Weby autorů

- mnozí autoři sdílí data na svých webech
- u článků vždy OA verze a odkaz na data
- většinou velmi kvalitně zpracovaná metadata
- často jediná cesta, jak se k datům dostat

# Rešerše dat | Data journals

- některé odborné časopisy umožňují přiložit data
- existují ale také speciální datové časopisy
- datové články – popis data a data
- někdy např. jako [data descriptor](#)
- i zde se provádí recenzní řízení
- mohou být i smíšená periodika



# Typy dat

- způsob uložení dat do souboru
- data.**XLSX**
- data.**CSV**, data.**TAB**
- data.JSON
- data.KML, data.GeoJSON – geografický rozměr
- .SAV, .DTA nebo .RDATA – různé analytické programy

# Vyhledávačka



Narazili jste na výzkum, který zkoumá obavy z terorismu spojené s přijímáním uprchlíků. Rádi byste si nad daty z výzkumu provedli **svoje vlastní analýzy** – zkuste původní dataset dohledat.

DOI článku: [10.31235/osf.io/d4ewg](https://doi.org/10.31235/osf.io/d4ewg)

# **Analýza dat**

- nadstavba prostého vyhledání dat
- vrstva práce s daty
- vyžaduje specifické dovednosti a znalosti

# Hodnocení kvality dat

- data nemusí být čistá – *GI-GO*
- data vznikají v kontextu, kontext je *messy*
- kdo je měří a publikuje? – *jasný zdroj!*
- jaká jsou k nim metadata?
- dokážu zjistit, co který sloupeček znamená?
- je popsána metodologie jejich vzniku?

# Čištění dat

- data nejsou vždy čistá
- především ta, která vznikají ručním zápisem
- existují nástroje, které zjednodušují jejich čištění
- práce pro datové knihovníky?



**OpenRefine**

# Citování dat

- **jak citovat dataset?**
- [existují různé návody](#)
- ISO 690 – jako záznam v repozitáři?
- specifikace on-line zdroje: [online databáze]
- mnohdy to autoři datasetu [přímo řeknou](#)