

Data v kyberprostoru

Každodennost v kyberprostoru

20. 3. 2024

Data

- výraz pro *údaje*, používané pro popis nějakého jevu
- popis vlastnosti pozorovaného objektu
- získávají se zápisem nebo měřením

KTD - Úplné zobrazení záznamu

Zvolte formát: [Standardní](#) -- [Termín & výklad](#) -- [Termín & angl.ekvivalent](#) -- [Heslo s historií](#)

Záznam 8 z 59

Termín	 <u>data</u>
Ekvivalent	údaje
Termín anglicky	data
Výklad termínu	<p>Reprezentace informací vhodně formalizovaná pro komunikaci, interpretaci a zpracování lidmi a automaty. Data mohou být reprezentována libovolnými řetězci znaků (čísel, příkazů, vět) uloženými na informačním nosiči. Data nemají zpravidla význam sama o sobě, ale teprve jsou-li pochopena, interpretována, komunikována a využita člověkem nebo počítačem, stávají se smysluplnými informacemi.</p> <p>[ODLIS] [STRAKA-1990:192] [OSVALDOVÁ,HALADA-1999:59] [ISO/FDIS 5127-2000:9]</p>
Normativní výklad	<p>Reprezentace informací formalizovaným způsobem vhodným pro komunikaci, interpretaci a zpracování [ISO 1087-2:2000]. [ČSN ISO 5127-2003]</p>
Zdroj výkladu	 ODLIS
	 STRAKA-1990:192
	 OSVALDOVÁ,HALADA-1999:59
	 ISO/FDIS 5127-2000:9
Zdroj norm.výkladu	 ČSN ISO 5127-2003

A DAY IN DATA

The exponential growth of data is undisputed, but the numbers behind this explosion – fuelled by internet of things and the use of connected devices – are hard to comprehend, particularly when looked at in the context of one day

500m

tweets are sent every day

Twitter



4PB

of data created by Facebook, including

350m photos

100m hours of video watch time

Facebook Research

DEMYSTIFYING DATA UNITS

From the more familiar 'bit' or 'megabyte', larger units of measurement are more frequently being used to explain the masses of data

Unit	Value	Size
b bit	0 or 1	1/8 of a byte
B byte	8 bits	1 byte
KB kilobyte	1,000 bytes	1,000 bytes
MB megabyte	1,000 ² bytes	1,000,000 bytes
GB gigabyte	1,000 ³ bytes	1,000,000,000 bytes
TB terabyte	1,000 ⁴ bytes	1,000,000,000,000 bytes
PB petabyte	1,000 ⁵ bytes	1,000,000,000,000,000 bytes
EB exabyte	1,000 ⁶ bytes	1,000,000,000,000,000,000 bytes
ZB zettabyte	1,000 ⁷ bytes	1,000,000,000,000,000,000,000 bytes
YB yottabyte	1,000 ⁸ bytes	1,000,000,000,000,000,000,000,000 bytes

*A lowercase 'b' is used as an abbreviation for bits, while an uppercase 'B' represents bytes.

463EB

of data will be created every day by 2025

IDC

294bn

billion emails are sent

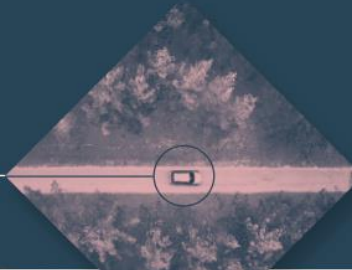
Radicati Group

320bn

emails to be sent each day by 2021

306bn

emails to be sent each day by 2020



4TB

of data produced by a connected car

Intel

65bn

messages sent over WhatsApp and two billion minutes of voice and video calls made

Facebook



95m

photos and videos are shared on Instagram

Instagram Business

3.9bn

people use emails



Smart Insights

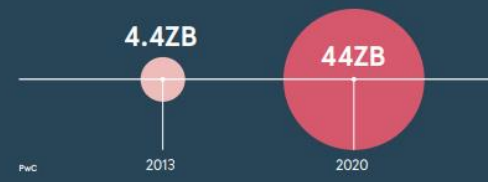
28PB

to be generated from wearable devices by 2020

Statista



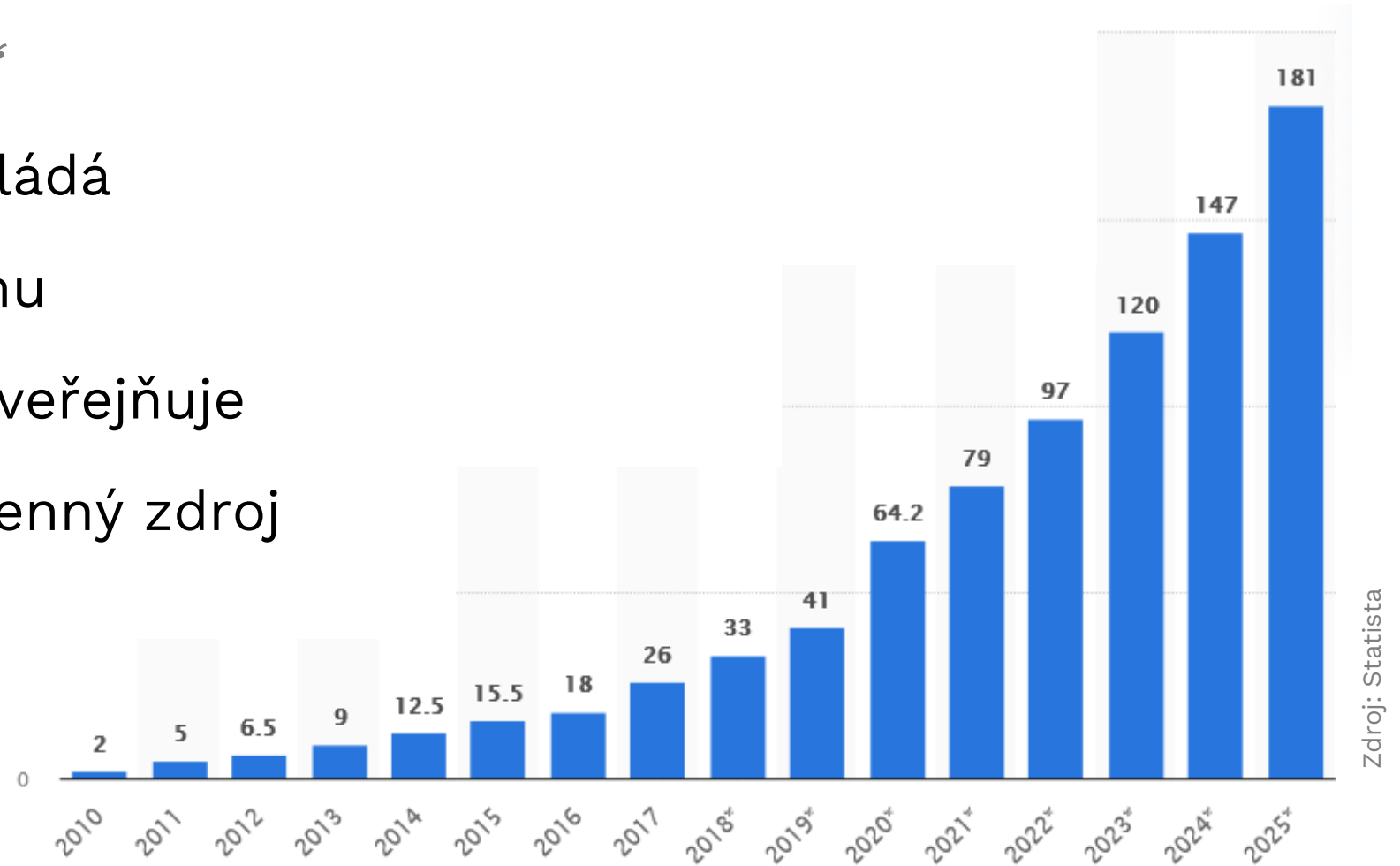
ACCUMULATED DIGITAL UNIVERSE OF DATA



PwC

Data kolem nás

- „data jsou nová ropa“
- všechno se měří a ukládá
- mnoho dat je k ničemu
- mnoho se jich dnes zveřejňuje
- některá představují cenný zdroj



data pro aktivní občanství

data v pracovním životě

datový mindset

data ve vědě

data pro řešení každodenních problémů

Evidence-based praxe

- EBM, EBL,...
- přístupy založené na datech
- využití dat pro rozhodování

- *capta*

Data v různých podobách

Různé podoby dat kolem nás...

- expertní systémy
- faktografické databáze
- statistické databáze
- datové repozitáře



I. Expertní systémy

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

I. Expertní systémy



Dobrý den, mám opakované bolesti hlavy,
zvracím a trápí mě kašel.
Co by mi mohlo být?



<https://symptoms.webmd.com/>

WebMD Symptom Checker BETA

Identify possible conditions and treatment related to your symptoms.

This tool does not provide medical advice. [See additional information.](#)

Age

Gender

Male

Female

Continue



INFO

SYMPTOMS

HISTORY

CONDITIONS

DETAILS

TREATMENT

What is your main symptom?

e.g., Headache

or Choose common symptoms

Bloating

Cough

Diarrhea

Dizziness

Fatigue

Fever

Headache

Muscle cramp

Nausea

Throat irritation

Gender Male

Age 29

My Symptoms

< Previous

Continue >

Conditions that match your symptoms

UNDERSTANDING YOUR RESULTS 

Tension headache 

Iron Deficiency Anemia 

Hypothyroidism (adult) 

Acute Sinusitis 

Drug allergy 



LOAD 10 MORE CONDITIONS

< Previous

Start Over

Tension Headache

Stress headache | Muscle spasm headache

CONDITION DETAILS

TREATMENT OPTIONS

Symptoms

Tension headache symptoms include dull pain and pressure, tightness like a band around the head, mild sensitivity to noise or light, difficulty sleeping, and difficulty concentrating.

How Common

Among U.S. adults, 30% to 80% have occasional tension headaches.

Overview

Tension headaches are the most common type of headache, caused by muscle tension in the neck, face, jaw, or shoulders. Many people have occasional tension headaches, which often start in the middle of the day. For others, tension headaches occur every day. Everyday stress, lack of sleep, ... [read more](#)

[Go to the Migraines & Headaches Health Center >](#)

II. Faktografické databáze

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

II. Faktografické databáze



Mám doma Citalec 20, ale ztratil jsem příbalový leták.
Nedá se nějak zjistit, jaké může mít vedlejší účinky?



<http://www.sukl.cz/modules/medication/search.php>

Faktografické databáze

Specializované FD

1. [Chemspider](#)
2. [Drugbank](#)
3. [CIA World Factbook](#)
4. [Featherbase](#)

A tisíce dalších...

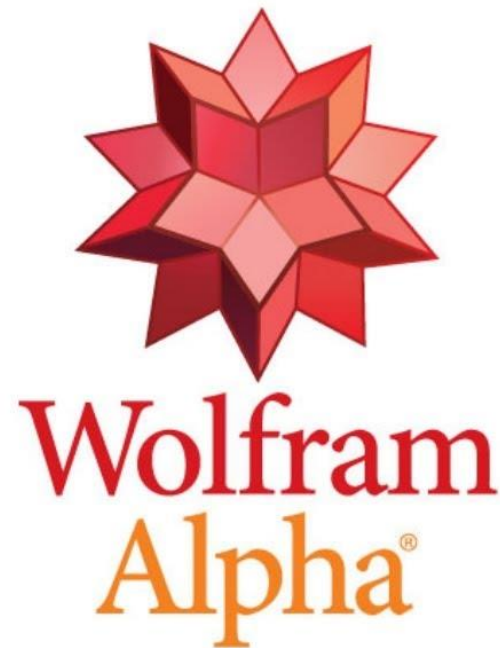
Pracovali jste někdy
s Wolfram Alpha?



II. Faktografické databáze

Wolfram Alpha

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



A screenshot of the Wolfram Alpha search results page. The search bar contains the text "International Space Station May 18 2009 6:00pm". Below the search bar, the "Input interpretation" section shows "International Space Station (spacecraft)" and "6:00:00 pm CDT | Monday, May 18, 2009". The "Position at 6:00 pm:" section features a world map with a red dot indicating the location of the International Space Station and a blue orbital path. The coordinates "22.51° North 94.35° West (ocean)" are displayed below the map. The Wolfram Alpha logo is visible at the top right of the page.

Wolfram Alpha | Tipy

- scrabble vs. jenga
- thorin vs. frodo
- magikarp vs. pikachu
- uncle's uncle's son's daughter's cousin
- libraries with number of books > 15 million
- 85 kg 192 cm 12 beers in 5 hours
- 42 mars bars
- F#

III. Statistické databáze

- obrovské množství databází mezinárodních organizací
- databáze od státních orgánů, úřadů
- statistické úřady

Proklikajte si:

<https://data.worldbank.org/>

<https://www.who.int/data/collections>

<https://data.unicef.org/>

<https://stats.oecd.org/>

<https://data.europa.eu/en>



III. Statistické databáze



Chci si otevřít hotel v Brně a potřeboval bych vědět, kolik se tady během roku ubytovává lidí a jestli to číslo klesá nebo stoupá – no prostě jestli má cenu další hotel otvírat...



<https://www.czso.cz/>

IV. Datové repozitáře

- repozitáře **otevřených dat**
- data ve vědě a vědeckém publikování
- institucionální repozitáře

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.

U nás je tato definice velmi důležitá...

Vaše město zatím
nesdílí svá data
otevřeně. Jakými
argumenty ho zkusíte
přesvědčit?



Open data | +/-

- zvýšení efektivity – sdílení a analýza
- podpora ekonomiky – zdroj inovací, surovina
- transparentnost a kontrola
- zapojení občanů do rozhodování
- datová žurnalistika

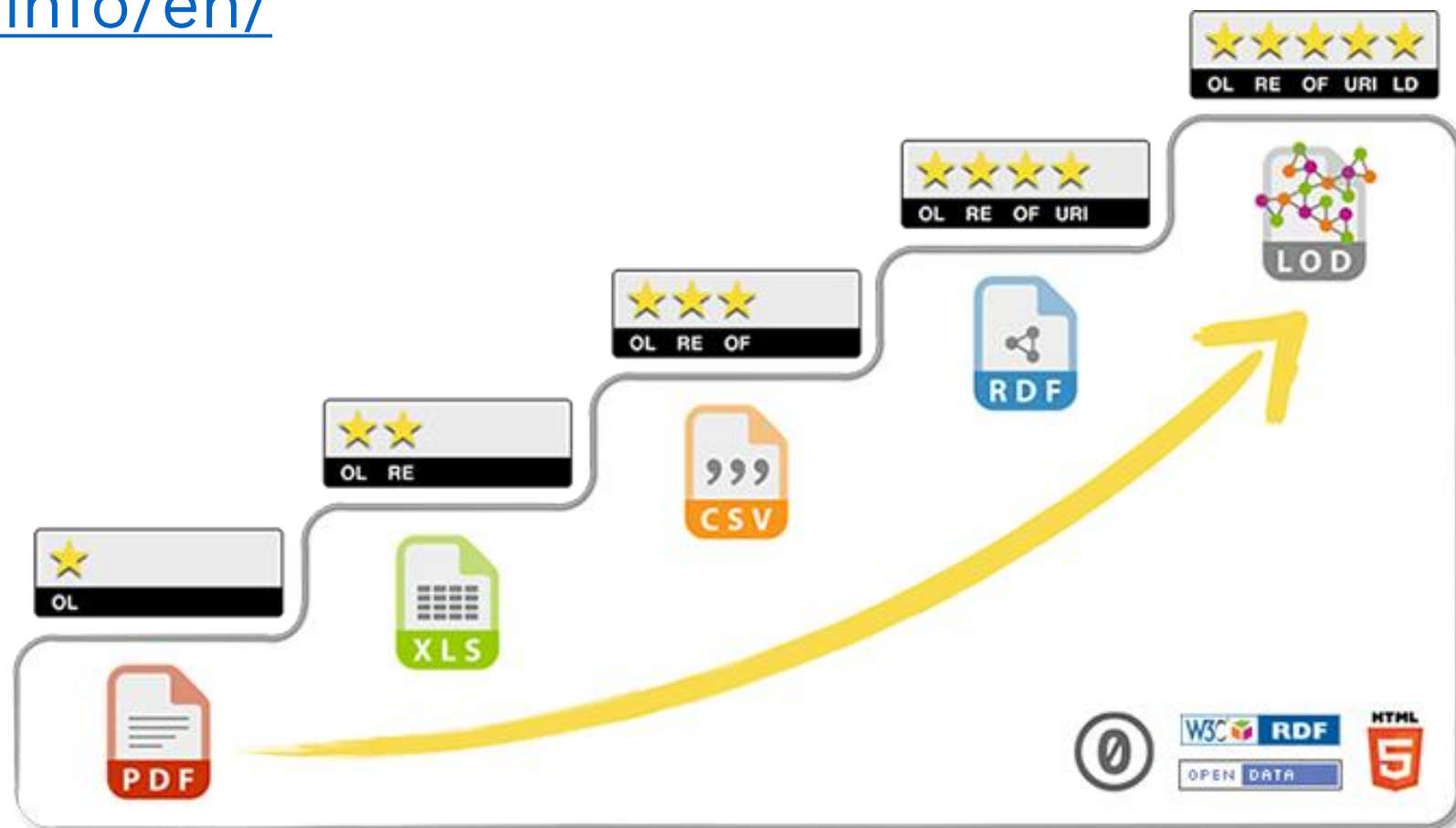


Open data | +/-

- vznik aplikací nad otevřenými daty
- různí „hlídači“ státu
- <https://www.znecistovatele.cz/>
- <https://hackujstat.cz/>
- najdete bezpochyby desítky dalších...

Open data | Stupně otevřenosti

<https://5stardata.info/en/>



Portál otevřených dat České republiky

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

135 881 datových sad

Název datové sady Hledat

Novinky ze světa otevřených dat

Webinář: Otevřená data ve zdravotnictví II

23. 4. 2021

V pátek 7. května 2021 v 10:30 pořádá Ústav zdravotnických informací a statistiky ČR [druhý webinář nad zdravotnickými daty](#). Cílem webináře je odpovědět na klíčové otázky o zdravotnických datech. Opět se můžete těšit na dva hosty, kteří zhodnotí aktuální stav otevřených dat a poskytnou zkušenosti s jejich zpracováním.

Webinář: Otevřená data ve zdravotnictví

23. 3. 2021

V pátek 26. března 2021 v 10:30 pořádá Ústav zdravotnických informací a statistiky ČR první online diskusní setkání nad zdravotnickými daty. Cílem webináře je odpovědět na aktuální otázky o zdravotnických datech. Pro účast na webináři je nutno se [registrovat](#).

Aktualizovali jsme vzdělávací sekci

5. 3. 2021

Rozšířili jsme náš [rozcestník vzdělávací sekce](#) a přidali aktualizovaný přehled online školení v oblasti otevřených dat. Intenzivně pracujeme na e-learningovém školení, které plánujeme spustit v září.

Archiv novinek »

File Edit View History Bookmarks Tools Help

data.europa.eu

https://data.europa.eu/en

data.europa.eu

The official portal for European data

English (en) Search site content Search

Discover the datasets from the former EU Open Data Portal [here](#).














Data Impact & Studies Training News & Events About

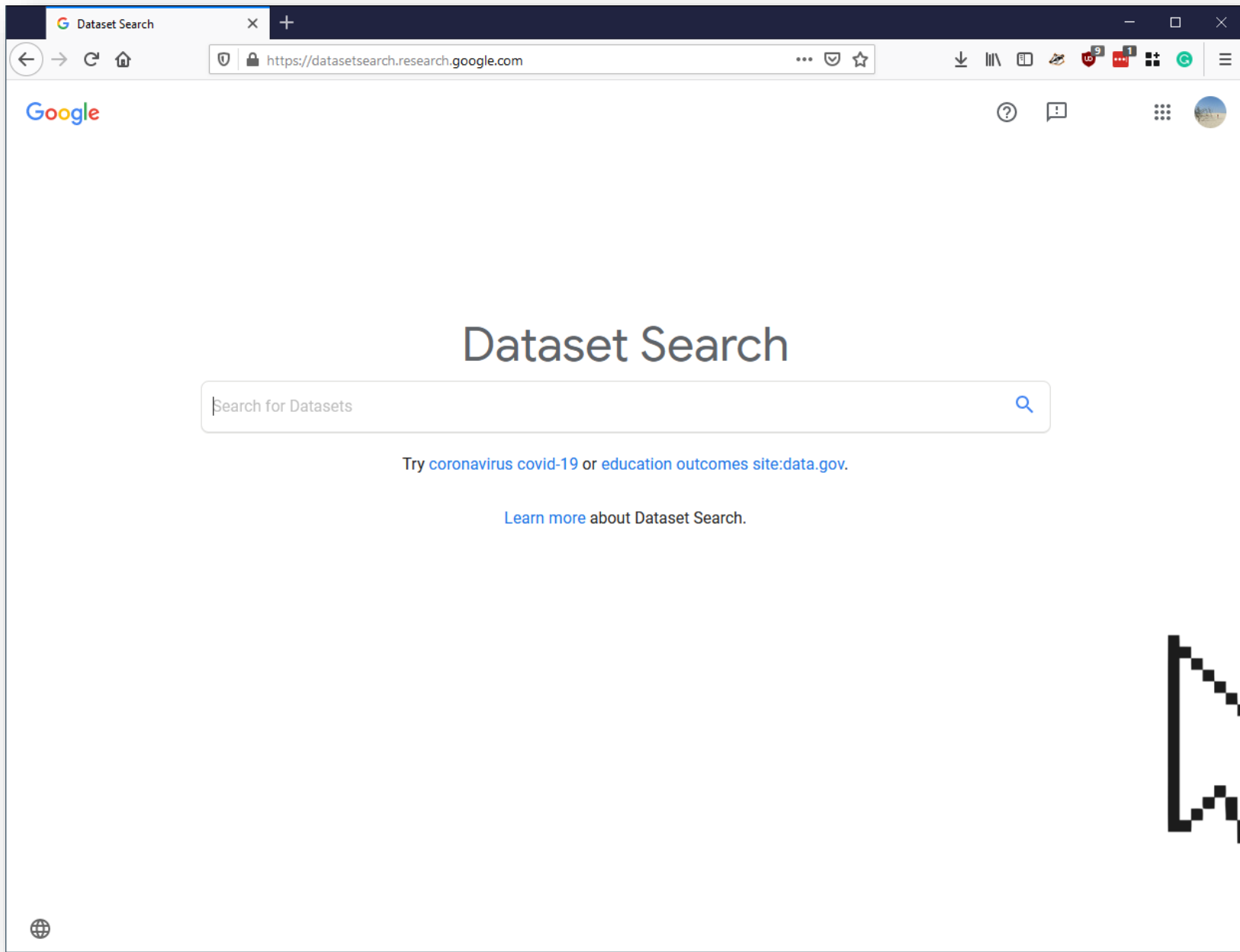
data.europa.eu

The official portal for European data

81 Catalogues **36** Countries **1125030** Datasets

Search datasets

 Agriculture, Fisheries, Forestry & Foods	 Economy & Finance	 Education, Culture & Sport	 Energy	 Environment	 Government & Public Sector
 Health	 International Issues	 Justice, Legal System & Public Safety	 Population & Society	 Regions & Cities	 Science & Technology
 Transport	<input type="text" value="Search datasets"/> Search				



Dataset Search

Try [coronavirus covid-19](#) or [education outcomes site:data.gov](#).

[Learn more](#) about Dataset Search.

IV. Datové repozitáře



Nedávno se u nás v obci stala nehoda, kdy se srazil vlak s autem. Chtěl bych o tom vědět víc, ale našel jsem jen pár zpráv v novinách a to je všechno.



<https://files.fpfis.tech.ec.europa.eu/index.php/s/A6Arm7TffA9d7yb>

Sdílení dat ve vědě

- reprodukovatelnost výzkumu
- ověření správnosti
- možnost navázat na předchozí výzkum
- jiné využití stejných dat
- sdílení prostředí a kontextu výzkumu
- [datové repozitáře](#)



Datové repozitáře

- [Zenodo](#)
 - [figshare](#)
 - [Dryad](#)
- obecně zaměřené*

Jak se vybírá?

- udělování identifikátoru
- práce s metadaty
- verzování datových souborů



Datové repozitáře

- *oborové* datové repozitáře
- re3data.org
- [LINDAT](https://lindat.cz)

- *národní* repozitáře
- <https://data.narodni-repozitar.cz/>
- <https://access.olos.swiss/portal/#/home>

Datové repozitáře

- institucionální repozitáře
- provozují např. univerzity
- <https://www.lib.cas.cz/asep/>
- <https://dataverse.org/>
- <https://dataverse.harvard.edu/>
- <https://ckan.org/>



Background image showing a blurred table of data with numerical values and commas.

Práce s daty

Typy dat

- *způsob uložení dat do souboru*
- XLSX, CSV, TXT
- JSON, XML
- KML, GeoJSON – geografický rozměr

CSV

	A	B	C	D
1	ID	Gender	City	Monthly_
2	ID000002C	Female	Delhi	20000
3	ID000004E	Male	Mumbai	35000
4	ID000007F	Male	Panchkula	22500
5	ID000008I	Male	Saharsa	35000
6	ID000009J	Male	Bengaluru	100000
7	ID000010K	Male	Bengaluru	45000
8	ID000011L	Female	Sindhudurg	70000
9	ID000012M	Male	Bengaluru	20000
10	ID000013N	Male	Kochi	75000
11	ID000014C	Female	Mumbai	30000
12	ID000016C	Male	Mumbai	25000
13	ID000018S	Female	Surat	25000
14	ID000019T	Female	Pune	24000
15	ID000021V	Male	Bhubaneswar	27000
16	ID000022V	Female	Howrah	28000

JSON

```
{  
  "Employee": [  
    {  
      "id": "1",  
      "Name": "Ankit",  
      "Sal": "1000",  
    },  
    {  
      "id": "2",  
      "Name": "Faizv",
```

```
<?xml version="1.0"?>
```

```
<contact-info>
```

```
<name>Ankit</name>
```

```
<company>Analytics Vidhya</company>
```

```
<phone>+9187654321</phone>
```

```
</contact-info>
```


Základní dovednosti

- datový mindset
- [otevřít CSV](#), XLSX a další běžné formáty
- převést do sloupců
- provést základní operace
- filtrovat

Pokročilejší práce s daty



Iniciativa „*Stop zvířím!*“ vyžaduje zákaz chovu zvířat ve městech. Ohrožují prý životy chodců a především cyklistů. Jak často se něco takového třeba v Brně děje, že by se cyklista srazil se zvířetem?



<https://data.brno.cz/>



Hodnocení kvality dat

- data nemusí být tzv. „čistá“
- data vznikají v kontextu: *kola vs. zvěř v Brně?*
- kdo je měří a publikuje? – jasný zdroj!
- jaká jsou k nim metadata?
- dokážu zjistit, co který sloupeček znamená?
- je popsána metodologie jejich vzniku?

Poslední zdroje Existující připojení data	Aktualizovat vše Dotazy a připojení Vlastnosti Upravit propojení Dotazy a připojení	Akcie Měny Zeměpis Datové typy	Seřadit Seřadit Seřadit Seřadit a filtrovat	Vymazat Použít znovu Upřesnit Datové nástroje	Dynamické doplňování Odebrat duplicitu Ověření dat Sloučit Relace Spravovat datový model
---	---	---	--	--	---

ATSK

MISTOSK	PRAH
Jižní spojka	Prah
Jižní spojka	Prah
Jižní spojka	Prah
Jižní spojka	Prah
Jižní spojka	Prah
Jižní spojka	Prah
Jižní spojka	Prah
Jižní spojka	Prah
Sokolská	Prah
Sokolská	Prah
Sokolská	Prah
Sokolská	Prah
Jižní spojka	Prah
Jižní spojka	Prah
Jižní spojka	Prah
Jižní spojka	Prah
Jižní spojka	Prah
Jižní spojka	Prah
Jižní spojka	Prah
Legerova	Prah
Legerova	Prah
Legerova	Prah
Legerova	Prah
Jižní spojka	Prah
Jižní spojka	Prah
Jižní spojka	Prah
Jižní spojka	Prah
Sokolská	Prah
Jižní spojka	Prah
Jižní spojka	Prah

Seřadit od A do Z
 Seřadit od Z do A
 Seřadit podle barvy
 Zobrazení listu
 Zrušit filtr z „MISTOSK“
 Filtrovat podle barvy
 Filtry textu
 Hledat
 5. května, sjezd Na Strži
 5. května, směr centrum,
 5. května, směr centrum, Praha 4
 5. května, ZC u domu č. p. 43
 5. května
 5. května 43
 5. května DC čp 1045
 5. května ZC čp 1007 Praha 4
 Nejsou zobrazeny všechny položky.
 OK Zrušit

Dotazy a připojení

Dotazy | Připojení
 Dotazů: 1
 7736a486-36b3-47c1-8fc4-65ef1ffabf61-mhmp_dopravni_prestupky_2016
 Počet načtených řádků: 647 038 Počet chyb: 76

Pokročilejší práce s daty

- dotazování DB
- parsing
- čištění dat
- georeferencing
- ...

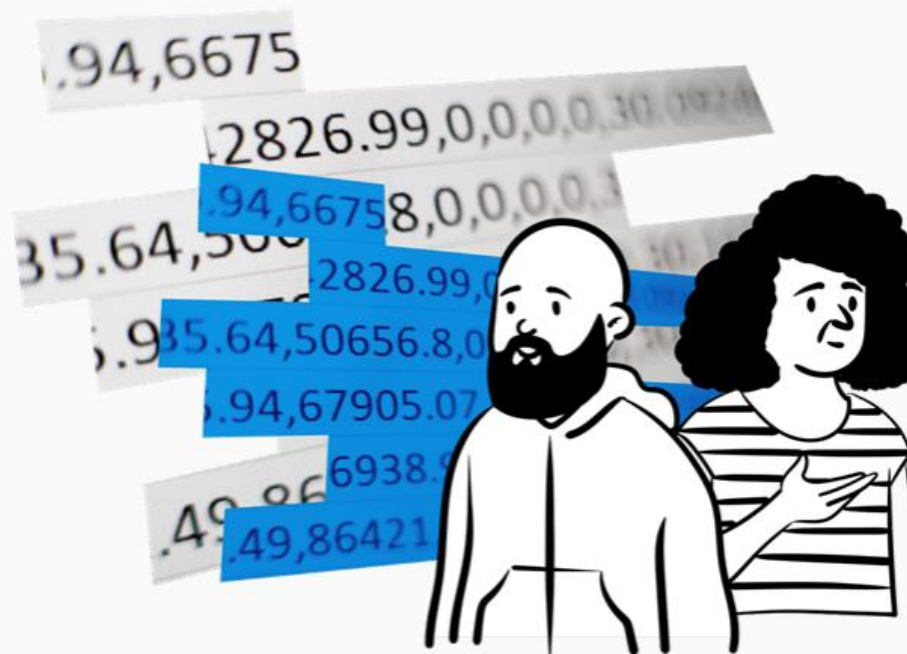
Datová gramotnost (pro knihovníky)

Mgr. Michal Koláček

Mgr. Tomáš Marek

Jako knihovníci dobře víme, že **žijeme ve světě dat**. Jejich smysluplné využití však vyžaduje specifické znalosti a dovednosti, které nám umožňují z datových zdrojů destilovat cenné informace, odpovědi a argumenty. V rámci kurzu otevřeného i začátečnickům si prakticky a názorně projdeme celým procesem práce s daty: naučíme se klást si nad nimi **ty správné otázky**, budeme **vyhledávat otevřená a volně dostupná data**, kriticky je **hodnotit, zkoumat, čistit a analyzovat**, abychom na závěr **mohli vytvořit vizualizaci a formulovat vlastní závěry**. Společně s lektory Michalem, Jáchymem a Tomášem si krok za krokem ukážeme i praktickou práci se skutečnými daty za využití základních nástrojů, jako je *Excel* nebo *Open Refine*.

Začít studovat





1 Datový mindset

1. Data a datový mindset

K čemu nám jsou data a jak můžeme v praxi aplikovat datový mindset? V prvním modulu se zaměříme na základy datové gramotnosti.



2 Zdroje dat

2. Zdroje dat

Ve druhém modulu se zaměříme na vyhledávání dat a odhalíme zajímavé a důvěryhodné zdroje, ze kterých můžeme data čerpat.



3 Metadata a import dat

3. Metadata a import dat

Ve třetím modulu se zaměříme na roli metadat a položíme si otázku, která nás bude provázet po celý zbytek kurzu.



4 První analýza

4. První analýza

V rámci první analýzy dat zkusíme odpovědět na úvodní otázku a využijeme k tomu například kontingenční tabulku.



5 Čištění dat

5. Čištění dat

Data nejsou nikdy úplně čistá a čím více volnosti panuje při jejich vzniku, tím jsou špinavější. V pátém modulu se podíváme, co s tím.



6 Transformace dat

6. Transformace dat

Abychom data dokázali správně analyzovat, bude si je vždy třeba připravit. V šestém modulu se nad našimi daty naučíme některé typy transformací.



7 Storytelling a vizualizace

7. Storytelling a vizualizace

Výstupy našich analýz chceme většinou někomu ukázat a efektivně komunikovat: na řadu tak přichází vizualizace a storytelling nad daty.



8 Závěrečný projekt

8. Závěrečný projekt

Nově získané dovednosti si v posledním modulu vyzkoušíme samostatně: dokážete pomocí základní analýzy dat odpovědět na zákludnou otázku?

Takže...

- dat je stále více a více
- částečně i díky trendu otevřených dat
- data hrají roli ve vědě i veřejné správě
- data přicházejí v různých podobách a formátech
- představují důležitý zdroj informací
- datový mindset

data pro aktivní občanství

data v pracovním životě

datový mindset

data ve vědě

data pro řešení každodenních problémů