

# CSV

**CSV** (z anglického **Comma-Separated Values**, česky **hodnoty oddělené čárkami**) je jednoduchý textový formát pro ukládání tabulkových dat, jako jsou čísla, text nebo datумы. Každý řádek souboru představuje **jeden záznam**, zatímco jednotlivé **hodnoty ve sloupcích** jsou odděleny **oddělovačem** (nejčastěji čárkou).

CSV je široce podporován tabulkovými procesory (např. Microsoft Excel, LibreOffice Calc), databázemi, programovacími jazyky (Python, JavaScript, PHP) a analytickými nástroji.

## Struktura souboru CSV

Základní princip:

- Každý **řádek** = jeden **záznam** (řádek tabulky).
- Každá **hodnota** ve sloupci je oddělena **znakem oddělovače**.
- První řádek často obsahuje **názvy sloupců** (tzv. hlavička, header).

Příklad souboru **zamestnanci.csv**:

```
jmeno,prijmeni,vek,oddeleni  
Jan,Novák,32,IT  
Eva,Svobodová,28,Marketing  
Petr,Dvořák,45,Finance
```

## Oddělovače

Přestože název „CSV“ znamená „oddělené čárkami“, v praxi se mohou používat i jiné znaky, zejména kvůli **lokalizaci** (např. v češtině se desetinná čísla zapisují s čárkou, nikoli tečkou):

- **Čárka (`,`)** - standard v anglicky mluvících zemích.
- **Středník (`;`)** - běžný v Evropě (včetně ČR) pro oddělení sloupců v Excelu.
- **Tabulátor (`\t`)** - pak se formát označuje jako **TSV** (Tab-Separated Values).

Příklad s oddělovačem středníkem:

```
jmeno;prijmeni;vek;oddeleni  
Jan;Novák;32;IT  
Eva;Svobodová;28;Marketing
```

## Pravidla pro uvozovky a speciální znaky

Pokud hodnota obsahuje:

- oddělovač (např. čárku),
- nový řádek,
- nebo uvozovky,

musí být **uzavřena do uvozovek** (obvykle dvojitých `„`).

Pravidla podle [RFC 4180](#) (standard pro CSV):

- Hodnoty s obsahem oddělovače, nového řádku nebo uvozovek musejí být uzavřeny v dvojitých uvozovkách.
- Uvnitř uvozovek se dvojitá uvozovka zapisuje jako dvě uvozovky: `“”`.

Příklad:

```
nazev,poznámka
"Projekt Alpha","Obsahuje čárku, teď je v pořádku"
"Kniha ""1984""","Klasika od George Orwella"
```

## Výhody a nevýhody

Výhody	Nevýhody
<b>Jednoduchost</b> – čitelné i pro lidi, snadno generovatelné.	<b>Žádná standardizace typů dat</b> – vše je text; čísla, datумы nebo logické hodnoty nejsou rozlišeny.
<b>Široká podpora</b> – většina software jej umí importovat/exportovat.	<b>Nejednoznačnost</b> – různé programy používají různé oddělovače, kódování, formáty datumů.
<b>Lehké soubory</b> – malá velikost, vhodné pro výměnu dat.	<b>Žádná podpora vztahů nebo struktury</b> – jen plochá tabulka, nelze reprezentovat hierarchii (na rozdíl od JSON nebo XML).
<b>Nezávislost na platformě</b> – funguje na Windows, Linux, macOS.	<b>Bezpečnostní rizika</b> – některé programy (např. Excel) automaticky interpretují obsah jako vzorce (tzv. <b>formula injection</b> ).

## Kódování znaků

Nejčastější kódování:

- **UTF-8** – doporučené pro mezinárodní výměnu, podporuje všechny znaky (včetně háčeků a čárek v češtině).
- **Windows-1250** – tradiční kódování pro češtinu ve Windows (dnes již méně doporučované).

□ Důležité: Pokud soubor obsahuje háčky a čárky (např. „Česká republika“), musí být uložen v kódování, které je podporuje (nejlépe **UTF-8 s BOM** pro kompatibilitu s Excelem na Windows).

## Použití v praxi

- **Export dat z databáze** – rychlý způsob, jak data přenést do Excelu.
- **Import produktových katalogů** – e-shopy často umožňují nahrát zboží přes CSV.
- **Zpracování dat v Pythonu** – modul `csv` umožňuje snadné čtení a zápis.

- **Migrační skripty** – přenos dat mezi systémy (např. z CRM do analytického nástroje).

Příklad v Pythonu:

```
import csv

with open('zamestnanci.csv', encoding='utf-8') as soubor:
    reader = csv.DictReader(soubor)
    for radek in reader:
        print(radek['jmeno'], radek['prijmeni'])
```

## Bezpečnostní doporučení

- Nikdy neotvírejte podezřelé CSV soubory v Excelu – mohou obsahovat **škodlivé vzorce** (např. `=cmd|' /C calc!A0`).
- Při importu do Excelu vždy zvol **„Text“** jako formát sloupce, ne „Obecný“.
- Ověř kódování a oddělovač před zpracováním.
- Nepoužívejte CSV pro citlivá data bez šifrování (např. osobní údaje podle GDPR).

## Standardy a normy

- **RFC 4180** – formální specifikace formátu CSV (včetně pravidel pro uvozovky, řádkování a MIME typ `text/csv`).
- **MIME typ:** `text/csv`
- **Přípona souboru:** `.csv`

## Související pojmy

- [TSV](#)
- [JSON](#)
- [XML](#)
- [Tabulkový procesor](#)
- [Databáze](#)
- [ETL](#)

## Externí odkazy

- RFC 4180 – Common Format and MIME Type for Comma-Separated Values (CSV) Files: <https://datatracker.ietf.org/doc/html/rfc4180>
- W3C – Model for Tabular Data and Metadata on the Web: <https://www.w3.org/TR/tabular-data-model/>

## Viz také

- [Import dat](#)
- [Export dat](#)
- [Kódování znaků](#)

From:

<https://serviceit.cz/> - **IT ENCYKLOPEDIÉ**

Permanent link:

<https://serviceit.cz/doku.php?id=csv>

Last update: **2025/12/31 22:04**

