

MUNI

# Závěrečná shrnující přednáška

CORE042: Data – odpověď na základní otázku života, vesmíru a vůbec...

13. přednáška

**Michal Růžička** <[ruzicka@ics.muni.cz](mailto:ruzicka@ics.muni.cz)> a kol.

Správa dat – Ústav výpočetní techniky MU

MUNI

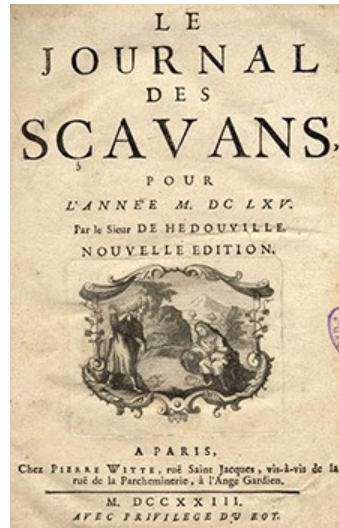
# O čem CORE042 byl

aneb Troška opakování na konec

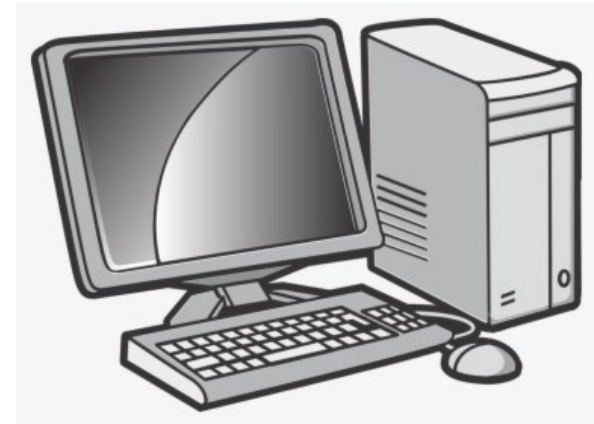
# Vědecká komunikace – historie a současnost



**osobní komunikace**



**tištěné časopisy**



**globální  
digitální  
komunikace**

# Životní cyklus výzkumných dat



Zdroj: ELIXIR RDMkit, <https://rdmkit.elixir-europe.org/>

- Jaká **data (znovu) používáte**
  - včetně licencí, které vám to umožňují,
- jaká **data generujete** a jakým způsobem,
- **kde je ukládáte, zálohujete, dlouhodobě uchováváte,**
- jak je **trvale a jedinečně identifikujete,**
- jak je **zpracováváte,**
- jak je **analyzujete,**
- kde je **zveřejňujete a sdílíte,**
- kdo tohle všechno **zaplatí;**
- **o čem data skutečně jsou,**
- **k čemu** jsou data **(ne)vhodná,**
- **kdo může data znovu použít,**
- jaká konkrétní data podporují vaše výsledky,
- jak je použít pro **opakování** vašich **experimentů** atd.

## Úvodní přednáška

### 1. přednáška

Michal Růžička <ruzicka@ics.muni.cz> a kol.

Správa dat – Ústav výpočetní techniky MU

### – Data ≠ informace!

– Data musí být správně pochopena, interpretována.

### – Data a nástroje a metody jejich zpracování jsou základem i výsledkem soudobého výzkumu a vědecké komunikace.

– Vyžadují náležitou péči a porozumnění.

Abychom o ně nepřišli, a to i v dlouhodobém horizontu.

Abychom správně designovali výzkum a interpretovali výsledky.

– Životní cyklus dat platí obecně napříč obory,

– ale implementace jeho fází je oborově závislá.

## Zpracování a analýza (velkých) dat

### 2. přednáška

Tomáš Rebok <rebok@ics.muni.cz>

Centrum CERIT-SC – Ústav výpočetní techniky MU

## Zpracování a analýza dat

### – Jak se na data dívat?

- Strukturovaná vs. nestrukturovaná vs. semi-strukturovaná.
- Výběr vhodného modelu pro zpracování a analýzu.

Důležitá je i znalost předpokládaných dotazů.

Tabulkové procesory, SQL databáze,  
NewSQL databáze, NoSQL databáze.

- Nebojte se být Big(Data). 😊

## Výpočetní a úložné infrastruktury v ČR

- Dostupné prostřednictvím e-INFRA CZ.  
CESNET, CERIT-SC, IT4I.
- Výpočetní a úložné kapacity pro náročné zpracování.  
Akademikům dostupné zdarma.
- Doplnkové služby pro podporu datového zpracování a analýzy.

## Rozhodnutí založená na datech: o myších a lidech

### 3. přednáška

Štěpán Mikula <[stepan.mikula@econ.muni.cz](mailto:stepan.mikula@econ.muni.cz)>

Katedra ekonomie – Ekonomicko-správní fakulta

### Důležité koncepty

- Kauzalitu může a nemusí provázet korelace, ale korelace nutně neznamená kauzalitu.
- Selection bias.
- Pro prokázání kauzality potřebujeme experiment.
- Pokud nelze provést experiment, lze za určitých podmínek identifikovat kauzální efekt s pomocí identifikačních strategií (DiD, RDD, ...).
- Proč nás zajímá kauzalita? Protože nám pomáhá činit správná a efektivní rozhodnutí.

## Jazykové korpusy

### 4. přednáška

Pavel Rychlý <pary@fi.muni.cz>,

Vít Suchomel <139723@mail.muni.cz>

Centrum zpracování přirozeného jazyka –  
Fakulta informatiky

### Souhrn

- Jazykové korpusy poskytují reprezentativní užití jazyka.
- Často vytvářené z webu.
- Velká data.
- Studenti/zaměstnanci MU mají přístup ke [Sketch Engine](#).



## Poučení z „krize důvěryhodnosti“ v psychologické vědě

5. přednáška

Hynek Cígler <[cigler@fss.muni.cz](mailto:cigler@fss.muni.cz)>

Katedra psychologie – Fakulta sociálních studií

### Take home message

- Replikační krize není známka, že věda nefunguje.  
Naopak jde o sebekorektivní mechanismus.
- Klíčovým požadavkem na dnešní vědu je transparence.
- A nejde jen o sdílení dat; jde o transparentci celé vědecké metody.

## Data v mikrobiologii

### 6. přednáška

Stanislava Bezdíček Králová

<kralova.s@sci.muni.cz> a kol.

Externista – Přírodovědecká fakulta MU

### Shrnutí

- Cyklus dat v mikrobiologii/biologii → využíváme všechny kroky.
- Generujeme velké množství variabilních dat (fotografie, HPLC, sekvence, měření, popisky, ...).
- Generujeme enormní množství metadat (GPS souřadnice, typy materiálů, množství materiálů, časy/doby odběrů, klinická data, asociovaná data, ...).
- Vysoká náročnost na **plan** → **collect** → **process** → **analyse**.
- Výrazný podíl **sharing + reusability**.
- Výrazný tlak na FAIR data (zejména **preserve + share** → **to be reused!**) .
  - Velké množství databází.
  - Specifické databáze.
  - Nevýhoda – databáze bez review procesu a kontroly → chyby.

## Data a vědecká komunikace

### 7. přednáška

Michal Lorenz <[lorenz@mail.muni.cz](mailto:lorenz@mail.muni.cz)>

Katedra informačních studií a knihovnictví –

Filozofická fakulta MU

### Význačné body

- Syrová data (raw data) – epistemický omyl, neexistující entita.
- Typy dat:
  - Kvantitativní, kvalitativní.
  - Strukturovaná, polostrukturovaná, nestrukturovaná.
  - Prvotní, doplňková, pomíjivá a derivovaná data.
  - Primární, sekundární a terciární data.
  - Indexikální, atributivní, meta- a paradata.
- Paradigmata vědy: Empirická → teoretická → průzkum dat.
- Trendy vědecké komunikace: E-science, gridové technologie, věda 2.0, repozitáře.
- Měření vědecké komunikace.

## Data pro vzdělávání – jak data správně sesbírat a interpretovat

### 8. přednáška

Lubomír Prokeš <[prokes@ped.muni.cz](mailto:prokes@ped.muni.cz)>

Katedra fyziky, chemie a odborné přípravy –

Pedagogická fakulta

### Význačné body

- Data a hlavní problémy systému českého školství:
  - Změny ve vzdělávání se nedaří dělat strategicky ani systémově.
  - Chybí dlouhodobá koncepce, systém je na různých úrovních roztržštěný a spolupráce vázne.
  - Ředitelé a učitelé škol jsou zavaleni byrokracií.
  - Školám chybí podpora a jasný směr.
  - Školství je dlouhodobě podfinancované.
    - Rozhodování založené na datech může pomoci.
- Vzdělávání založené na důkazech.
- Metody sběru dat.
- Přehled pastí a propastí sběru a analýzy dat.

## Etické a právne aspekty práce s dátami

### 9. přednáška

Lukáš Hamřík <[lukas.hamrik@law.muni.cz](mailto:lukas.hamrik@law.muni.cz)>  
Ústav pro otázky soudnictví – Právnická fakulta

### Zhrnutie

- Etika a integrita vedeckej práce ako základný rámec academickej práce.
  - Prítomné v každej fáze výskumu.
  - Je potrebné rešpektovať a oboznámiť sa s pravidlami vedeckej práce.
- Právne otázky práce s dátami a vo výskume:
  - Minimalizmus pri práci s osobnými údajmi.
  - Pri práci s databázami – opatrnosť (autorské práva).
- Životný cyklus dát = neustále hľadanie rovnováhy medzi etikou, právom, efektivitou, transparentnosťou, atď.

## Otevřená data ve veřejné správě a EOSC v ČR

### 10. přednáška

Martin Dvořák <dvorak@ics.muni.cz>

EOSC-CZ Sekretariát – Ústav výpočetní techniky MU

### Shrnutí

#### Otevřená data ve veřejné správě

- Legislativně ukotvená.
- Bottom up i top down přístupy pro zveřejňování.
- Dobré postavení ČR v EU.
- Míra adopce.
- Velké rozdíly mezi poskytovateli v přístupu i datové kvalitě.
- OD jako základ digitalizace státu.
- data.Brno.

#### EOSC a FAIR data

- Není legislativně ukotveno. (!)
- Iniciativa OS existuje mnoho let, nyní se v ČR buduje infrastruktura.
- Definice “rules of the game”.
- Velmi mnoho aktérů.
- Napojení na evropský EOSC?
- Napojení na OD veřejné správy?

## Od akademického výzkumu k praxi – komercializace vědeckých výsledků

11. přednáška

Matěj Bůril <[buril@ctt.muni.cz](mailto:buril@ctt.muni.cz)>

Centrum transferu technologií (CTT MU)

### Motto

- I věda může být business a je to dobře.
- Aby věda byla k něčemu, musí najít své praktické uplatnění, ideálně pak ve formě reálného produktu nebo služby...

## Základy umělé inteligence

### 12. přednáška

**Michal Bozděch**

[<michal.bozdech@fsps.muni.cz>](mailto:michal.bozdech@fsps.muni.cz)

Katedra tělesné výchovy a společenských věd –  
Fakulta sportovních studií

### Význačné body

- Metody strojové inteligence můžete se správným software aplikovat i bez hlubších znalostí programování.
- Typy strojového učení:
  - Supervised learning – učení pod dozorem.
  - Unsupervised learning – učení bez dozoru.
  - Reinforcement learning – zpětnovazební učení.
- Ukázky aplikací, např. pro sportovní trénink.



# Ještě dva tipy na závěr

- [Důvěryhodnost vědy – podcast Leonardo Plus ČRo](#)

- Proč někteří odmítali roušky a očkování? Selhala důvěra veřejnosti v politiky, soudí bioetik Marek Orko Vácha.

- [Science Café v LMC : THEaiTRE – umělá inteligence jako dramatik](#)

- Pěkná přednáška o AI ChatGPT-2.

- Srozumitelně vysvětluje principy a upozorňuje na limity AI systémů specificky i obecně.

**MUNI**

# **Zadání témat závěrečných odborných písemných prací**

# Co jsme po vás chtěli

- Abychom se fyzicky vídali na přednáškách. 😊
- Abyste se **nebáli** na cokoliv **zeptat**.
  - Klidně v průběhu přednášky otevřete diskusi s přednášejícím a ostatními posluchači.  
Může podléhat změnám dle potřeb konkrétního přednášejícího. 😊
- Abyste nám **dali zpětnou vazbu** na kurz.
  - Co se vám **líbilo**.
  - Co se vám **nelíbilo**.
  - Co byste doporučili **dělat jinak**.
  - Buďte **upřímní**, otevření, **konstruktivní**.
  - CORE042 je **mladý kurz**, zpětná vazba od vás posluchačů je pro nás **nesmírně důležitá**.  
Kdykoliv během semestru, určitě na konci semestru.  
Mailem, přes Teamsy, osobně, v předmětové anketě v ISu, ...
- Abyste **na závěr odevzdali krátký odborný text** na zvolené téma.

# Závěrečný odborný text

- **3 kredity** za absolvování CORE042 **nedostanete úplně zadarmo.** 😊
  - **Nebudeme** zkoušet naučené **konkrétní informace**, **technické znalosti** apod.
  - Chceme, **abyste se** nad tématy **zamysleli**.
- Na konci semestru **odevzdáte krátký odborný text** na některé ze zadaných témat.
  - Krátký = **1–2 strany A4** (≈ 6–9 tisíc znaků včetně mezer) + **užité zdroje**.

Někdo to po vás všech musí také přečíst, takže za větší rozsah nejsou kladné body navíc. 😊
  - Odborný = důvěryhodné **zdroje**, **citace**, ...
  - **Odevzdat do Ne 2. 6. 2024** (tj. do ~3 týdnů od zadání témat na poslední přednášce).

Abychom práce do konce zkouškového období zvládli ohodnotit.
  - K výběru z 44 témat navazujících na přednášky, které v průběhu semestru uslyšíte.
- Závěrečné **hodnocení** jako kolokvium, tj. **binární: prospěl/neprospěl**

## – Odevzdání:

- Do zhruba 3 týdnů, tj. **do Ne 2. 6. 2024 23:59 CEST**.
- Formát **PDF**.
- Rozsah **1–2 strany A4** (≈ 6–9 tisíc znaků včetně mezer) + **užité zdroje**.
- **Do Odevzdáárny** závěrečných prací předmětu CORE042 v IS MU:  
<https://is.muni.cz/auth/el/cus/jaro2024/CORE042/ode/esej/>

Odevzdáárny vázané na balíky témat, ke kterým se přihlásíte.

## Úvodní přednáška

### 1. přednáška

Michal Růžička <[ruzicka@ics.muni.cz](mailto:ruzicka@ics.muni.cz)> a kol.  
Správa dat – Ústav výpočetní techniky MU

### Témata pro závěrečnou práci:

1. Stačí jen sdílet výzkumná data? Jak zabránit jejich nepochopení nebo dezinterpretaci?
2. Stinné stránky digitalizace vědecké komunikace – co nám digitalizace dat a zahlcení globální digitální komunikací ve vědě vzalo?
3. Doba digitálního temna – přežijí nás naše data?

## Zpracování a analýza (velkých) dat

### 2. přednáška

Tomáš Rebok <[rebok@ics.muni.cz](mailto:rebok@ics.muni.cz)>

Centrum CERIT-SC – Ústav výpočetní techniky MU

### Témata pro závěrečnou práci:

1. Veřejnost jako zdroj Big Data pro výzkum? Jak by se dala využít spolupráce s občany pro sběr dat využitelných pro výzkum?

## Rozhodnutí založená na datech: o myších a lidech

### 3. přednáška

Štěpán Mikula <[stepan.mikula@econ.muni.cz](mailto:stepan.mikula@econ.muni.cz)>

Katedra ekonomie – Ekonomicko-správní fakulta

### Témata pro závěrečnou práci:

1. Jaké dopady může mít záměna korelace za kauzalitu na tvorbu efektivních veřejných politik?
2. Náhoda nám pomáhá identifikovat kauzální vztahy. Proč tomu tak je?
3. Řízený experiment je nejjistější způsob, jak identifikovat kauzalitu. Někdy však není možné experiment z technických nebo etických důvodů provést. Je to konec? Je možné identifikovat kauzalitu i na základě primárně neexperimentálních dat?

## Jazykové korpusy

### 4. přednáška

Pavel Rychlý <pary@fi.muni.cz>,

Vít Suchomel <139723@mail.muni.cz>

Centrum zpracování přirozeného jazyka –  
Fakulta informatiky

### Témata pro závěrečnou práci:

1. Poskytují jazykové korpusy opravdu reprezentativní pohled na užití jazyka? Většina korpusů pochází z veřejných textů – co ale mluvený jazyk, soukromá textová komunikace nezachycená ve sklízených veřejných textech, ...?
2. Je anglická Wikipedia důvěryhodný zdroj informací?
3. Jak mohu využít korpus (např. v systému Sketch Engine) při psaní odborných článků v angličtině?



## Poučení z „krize důvěryhodnosti“ v psychologické vědě

### 5. přednáška

Hynek Cígler <[cigler@fss.muni.cz](mailto:cigler@fss.muni.cz)>

Katedra psychologie – Fakulta sociálních studií

### Témata pro závěrečnou práci:

1. Je replikační krize hrozbou pro důvěru veřejnosti ve vědu? Pro zpracování tématu vyhledejte a použijte dostupné odborné zdroje či komentáře.
2. Zamyslete se nad replikovatelností výzkumu v oboru, který studujete. Existují informace o míře replikovatelnosti? Probíhá nebo je reflektována krize zobecnitelnosti ve vašem oboru? Úvahu opřete o relevantní odbornou literaturu.
3. Pro studující, kteří plánují, realizují či zamýšlí realizaci empirického výzkumu (v rámci diplomové práce, seminární práce či třeba jen jako projekt během výuky): Na vámi vybraném tématu ilustруйте rozdíly mezi reprodukovatelností, replikovatelností a zobecnitelností. Jaké kroky byste mohli pro zajištění těchto tří atributů uskutečnit?

## Data v mikrobiologii

### 6. přednáška

**Stanislava Bezdíček Králová**

**<kralova.s@sci.muni.cz> a kol.**

Externista – Přírodovědecká fakulta MU

### Témata pro závěrečnou práci:

1. Co si představíte pod pojmem metadata (i vzhledem k vašemu oboru), jak byste si navrhli jejich organizaci a ukládání?
2. Jaké možnosti uchování celkových výzkumných dat si umíte představit?
3. Když si představíte zpracovávání biologických dat (nebo dat z vašeho oboru) v minulosti a v současnosti, jaké změny vidíte a čím jsou způsobeny?
4. Digitální laboratorní deníky – výhoda nebo komplikace proti sešitu s propiskou na laboratorním stole?

## Data a vědecká komunikace

### 7. přednáška

**Michal Lorenz** <[lorenz@mail.muni.cz](mailto:lorenz@mail.muni.cz)>

Katedra informačních studií a knihovnictví –

Filozofická fakulta MU

### Témata pro závěrečnou práci:

1. Ticho jak prostředek komunikace? Nicota a prázdno jako důležitá data?
2. Proč neexistují raw data?
3. Data, informace, znalost, moudrost... Kde jsou hranice?
4. Jak zajistit důvěryhodnost svého výzkumu, když nemohu zpřístupnit primární data?
5. Proč jsou metadata užitečný až nezbytný doplněk dat?
6. Citizen science – jak do výzkumu zapojit běžné občany?
7. Umělecká tvorba a AI – komu náleží autorská práva a odměny?
8. Existuje spravedlivý způsob, jak měřit vědecký výkon a kvalitu výzkumu?

## Data pro vzdělávání – jak data správně sesbírat a interpretovat

### 8. přednáška

Lubomír Prokeš <[prokes@ped.muni.cz](mailto:prokes@ped.muni.cz)>

Katedra fyziky, chemie a odborné přípravy –  
Pedagogická fakulta

### Témata pro závěrečnou práci:

1. Jak skloubit data a emoce – daty podložené rozhodování vs. jednání dle citu, kde přináší výhody, kde způsobuje problémy?
2. Data jako základ kritického myšlení.
3. Moje zkušenost s platformami pro vzdělávání využívajících analýzu Big Data – Duolingo apod.
4. Jak se v praxi vyhnout kognitivnímu zkreslení?
5. Při interpretaci dat se dá udělat řada chyb, viz některé jevy zmíněné na přednášce, jsou studenti v ČR od základních škol po univerzity vhodně vzdělávání v práci s daty („datová gramotnost“) a v kritickém myšlení? Co by se dalo zlepšit? Vaše osobní zkušenosti?

## Etické a právní aspekty práce s datami

### 9. přednáška

Lukáš Hamřík <[lukas.hamrik@law.muni.cz](mailto:lukas.hamrik@law.muni.cz)>

Ústav pro otázky soudnictví – Právnická fakulta

### Témata pro závěrečnou práci:

1. Je možné bránit' sa proti predátorským časopisom? Ak áno, tak ako?
2. Existuje jednoznačná hranica medzi etikou a efektivitou vo výskume? Teda medzi právami respondentov/účastníkmi pozorovaní a limitovanými zdrojmi výskumníkov (čas, financie, ...)?

## Otevřená data ve veřejné správě a EOSC v ČR

10. přednáška

Martin Dvořák <dvorak@ics.muni.cz>

EOSC-CZ Sekretariát – Ústav výpočetní techniky MU

### Témata pro závěrečnou práci:

1. Proč by měla být všechna data veřejné správy otevřená “by default”? A proč ne?
2. Měla by být FAIR vědecká data součástí otevřených dat veřejné správy?
3. Jaký je rozdíl v datovém managementu města Brna a města Prahy, zejména z hlediska politiky otevřených dat.

## Od akademického výzkumu k praxi – komercializace vědeckých výsledků

11. přednáška

Matěj Bůril <[buril@ctt.muni.cz](mailto:buril@ctt.muni.cz)>

Centrum transferu technologií (CTT MU)

### Témata pro závěrečnou práci:

1. Open science a komercializace duševního vlastnictví – kamarádi, anebo rivalové?
2. Uvolnění patentové ochrany ve vztahu k pandemii COVID-19 – ano či ne?
3. Jak prakticky podpořit zakládání akademických spin-off společností v ČR?
4. Patentová ochrana – motor nebo brzda pokroku?

## Základy umělé inteligence

### 12. přednáška

**Michal Bozděch**

[<michal.bozdech@fsps.muni.cz>](mailto:michal.bozdech@fsps.muni.cz)

Katedra tělesné výchovy a společenských věd –  
Fakulta sportovních studií

### Témata pro závěrečnou práci:

1. Etické rozměry rozhodování umělé inteligence: Proč je důležité porozumět mechanismům rozhodování AI?
2. Skrytá přítomnost: Jak a kde se umělá inteligence projevuje v našem každodenním životě?
3. Meze lidského a strojového myšlení: Mohou stroje nahradit lidský rozum a rozhodování?
4. Umělá inteligence v omezených prostředích: Kde je použití AI obtížné nebo nevhodné?
5. Den v životě s umělou inteligencí: Jak se setkáváme s různými typy AI během našeho každodenního života?



# Závěrečná písemná práce – přihlašování

- Všechna témata jsou vypsána formou balíků témat předmětu:

<https://is.muni.cz/auth/ro/cus/jaro2024/CORE042/esej>

- **Přihlašovat se může zde:**

<https://is.muni.cz/auth/ro/cus/jaro2024/CORE042/esej/stud>

- **Přihlašování** probíhá **od konce dnešní přednášky**, tj. 14. 5. 2024 od 14:00.

- **Limit** počtu přihlášených pro každé téma jsou **3 studenti**.

Máme celkem 44 témat, pro 105 zapsaných studentů to představuje rozumnou volnost ve výběru tématu při zachování dostatečné rozmanitosti zpracovaných témat.

- Každý **student odevzdává** písemnou práci na **téma**, ke kterému je **přihlášen!**

MUNI

# Jak se vám CORE042 líbil?

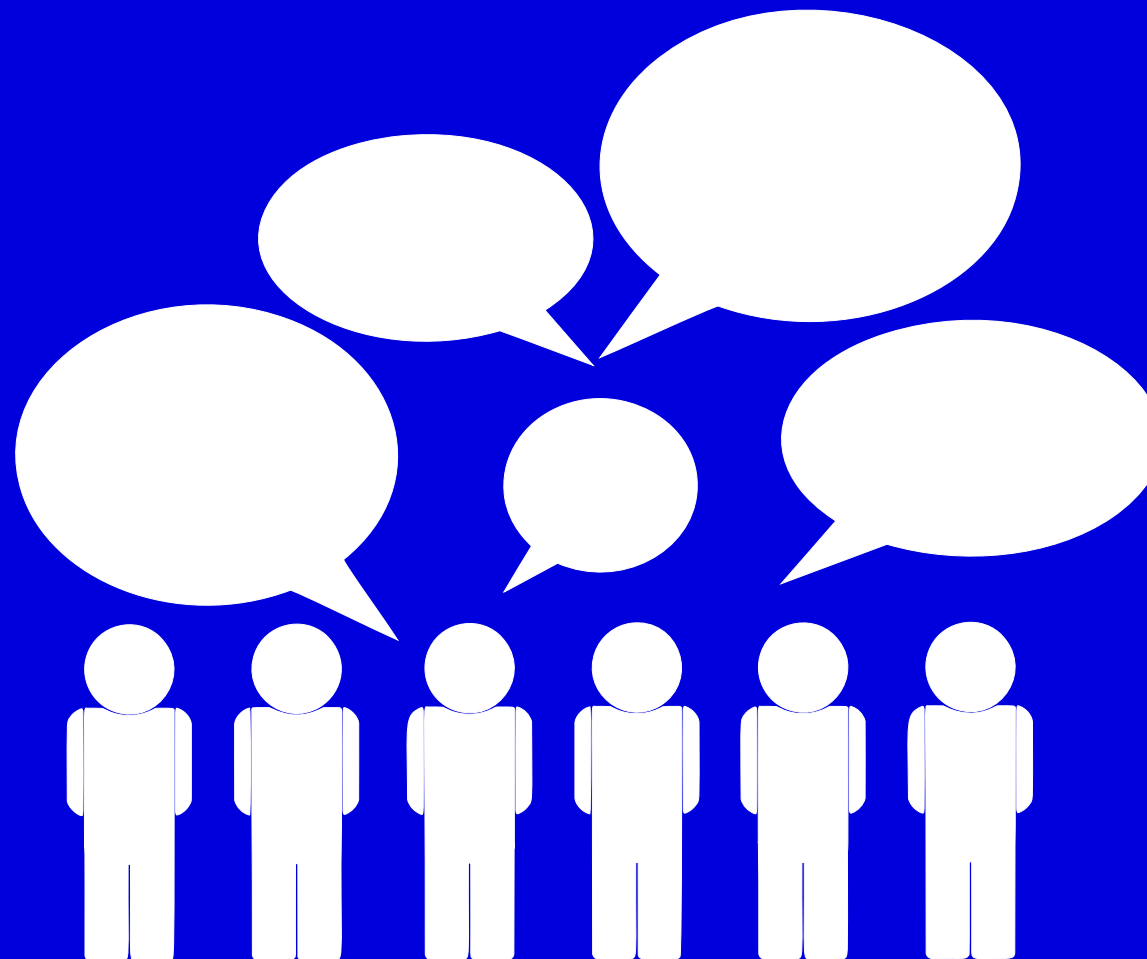
Potřebujeme zpětnou vazbu od vás všech!

# Zpětná vazba, jak kurz CORE042 zlepšit pro příští běh

- Řekněte nám to **do očí teď a tady!** 😊
- Vyplňte prosím **předmětovou anketu**, napište nám mail, přes Teams, diskusní fórum v IS MU, ...
- **Vaše nápady na zlepšení CORE042 pro další běh:**
  - Říkejte.
  - Pište na [core42@space.muni.cz](mailto:core42@space.muni.cz).

# MUNI

## Diskuse



Zdroj: [Communicate\\_communication\\_conference\\_2028004](#) od [OpenClipart-Vectors](#) z [Pixabay](#)