

**MUNI**

# **Závěrečná shrnující přednáška**

CORE042 – Výzkum v 21. století: data pohledem vaším a vašich kolegů  
13. přednáška

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

MUNI

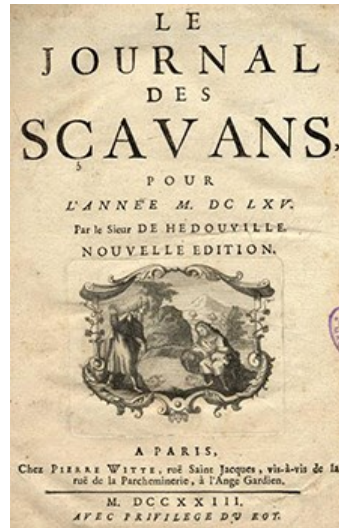
# O čem CORE042 byl

aneb Troška opakování na konec

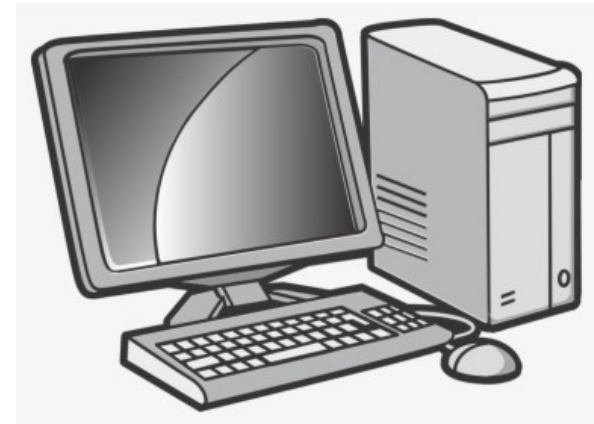
# Vědecká komunikace – historie a současnost



**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.

Bezpečnost a 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>

Divize IT infrastruktury – 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
- doplňkové 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>

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 PŘF – Ústav experimentální biologie 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.

# MUNI

## 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.

## Od akademického výzkumu k praxi – komercializace vědeckých výsledků

10. 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...

## Vědecký proces

### 11. přednáška

**Pavel Tomančák**

<pavel.tomancak@ceitec.cz>

Vědecká rada – Středoevropský technologický institut  
(CEITEC)

### Vědecký proces

- Položit si správnou vědeckou otázku.
- Shromáždit odpovídající tým a získat finanční prostředky na výzkum.
- Udělat vědu.
- Publikovat výsledky.
- Pozorovat impakt objevů a odrazit se od nich k dalším otázkám.

### Význačné body

- Dlouhodobě udržovat data je komplikované, pamatujte na *bus factor*. 😊
  - Používejte oborové repozitáře – situace se stále zlepšuje.
- Problémy s dostupností a věrohodností dat jsou starší než počítačové zpracování.
- Důležitá jsou nejen data, ale také celé prostředí, *pipelines*, které je zpracovávají.
  - Potřebné pro replikaci výzkumu, znovuvyužitelnost, ...

## Základy umělé inteligence

### 12. přednáška

**Michal Bozděch**

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

Katedra kineziologie – 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.
- Typy strojového učení.
- Ukázky aplikací, např. pro sportovní trénink.



MUNI

# Zadání závěrečných esejí

# 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 **zcela nový 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 odevzdali závěrečnou esej.**

# Závěrečná esej

- **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 krátkou esej** na některé ze zadaných témat.
  - Krátká = **1–2 strany A4** (≈ 6–9 tisíc znaků včetně mezer)
    - Někdo to po vás všech musí také přečíst, takže za větší rozsah nejsou kladné body navíc. 😊
    - Velikost písma, řádkování a rozložení stránky musí být „rozumné“.
  - **Odevzdat do konce roku**, abychom během zkouškového období ohodnotili.
  - Akceptujeme i jiné závěry a **argumentací podložená** zpochybnění našich tvrzení.
  - Recese pouze v rozumné míře.
  - Témata esejí budou navazovat na přednášky, které v průběhu kurzu uslyšíte.
- Závěrečné **hodnocení** jako kolokvium, tj. **binární: prospěl/neprospěl**

## – Odevzdání:

- Do konce roku, tj. **do 31. 12. 2022 23:59 CET.**
- Formát **PDF**.
- Rozsah **1–2 strany A4** (≈ 6–9 tisíc znaků včetně mezer).
- **Do Odevzdáárny** předmětu CORE042 v IS MU:  
[https://is.muni.cz/auth/el/cus/podzim2022/CORE042/ode/ode\\_zaverecna\\_esej/](https://is.muni.cz/auth/el/cus/podzim2022/CORE042/ode/ode_zaverecna_esej/)

## Úvodní přednáška

### 1. přednáška

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

Bezpečnost a správa dat – Ústav výpočetní techniky MU

### Témata pro závěrečnou esej:

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)>

Divize IT infrastruktury – Ústav výpočetní techniky MU

### Témata pro závěrečnou esej:

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 esej:

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>

Centrum zpracování přirozeného jazyka – Fakulta  
informatiky

### Témata pro závěrečnou esej:

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 sklizených veřejných textech, ...?
2. Je anglická Wikipedia důvěryhodný zdroj informací?

## 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 esej:

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?
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.



## Data v mikrobiologii

### 6. přednáška

**Stanislava Bezdíček Králová**

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

Externista PřF – Ústav experimentální biologie MU

### Témata pro závěrečnou esej:

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 esej:

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 esej:

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 esej:

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, ...)?

## Od akademického výzkumu k praxi – komercializace vědeckých výsledků

10. přednáška

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

Centrum transferu technologií (CTT MU)

### Témata pro závěrečnou esej:

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?

## Vědecký proces

11. přednáška

**Pavel Tomančák**

[<pavel.tomancak@ceitec.cz>](mailto:pavel.tomancak@ceitec.cz)

Vědecká rada – Středoevropský technologický institut  
(CEITEC)

### **Témata pro závěrečnou esej:**

1. Vědecké podvody s „daty“ a vzorky před érou počítačů.

## Základy umělé inteligence

### 12. přednáška

**Michal Bozděch**

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

Katedra kineziologie – Fakulta sportovních studií

### Témata pro závěrečnou esej:

1. Proč je důležité rozumět důvodům rozhodování umělé inteligence?
2. Kde všude se můžeme setkat s umělou inteligencí, aniž by to bylo na první pohled patrné?
3. Je možné, aby umělá inteligence nahradila lidský rozum a rozhodování?
4. Ve kterém prostředí, či v jaké situaci není možné nebo je nevhodné použití umělou inteligencí?
5. Popište svůj běžný den, ve kterém přijdete do kontaktu s co nejvíce typy umělé inteligence.
6. Pozitivní a negativní důsledky přílišného používání umělé inteligence v běžném světě.

**MUNI**

# **Jak se vám CORE042 líbil?**

Potřebujeme zpětnou vazbu od vás všech!



# Zpětná vazba pro MU na CORE předměty

- Řekněte nám to **do očí teď a tady!** 😊
- Vyplňte prosím **předmětovou anketu.**
- Vyplňte **anketu specifickou pro CORE předměty.**
  - Bude zasílat rektorát MU nad rámec běžné předmětové ankety.

*Výuka předmětů CORE je novým prvkem v životě studujících i vyučujících, je průzkum k výuce s tímto dotazníkem prováděn navíc k běžné předmětové anketě, kterou nenahrazuje, nýbrž rozšiřuje a doplňuje.*

*Zvláštní dotazník obsahuje otázky k výuce, na které v předmětové anketě není prostor.*

*Prvořadým cílem dotazníku je získat informace k reflexi výuky ze strany studujících.*

*Dotazník je pro studující **naprosto anonymní.***

*Hned po Novém roce, 2. ledna 2023, dostanou studenti CORE kurzů do e-mailových schránek dopis s informacemi k tomuto průzkumu a s odkazem, pod nímž budou moci dotazník vyplnit.*

# Jak kurz CORE042 zlepšit pro příští běh

## – Naše nápady:

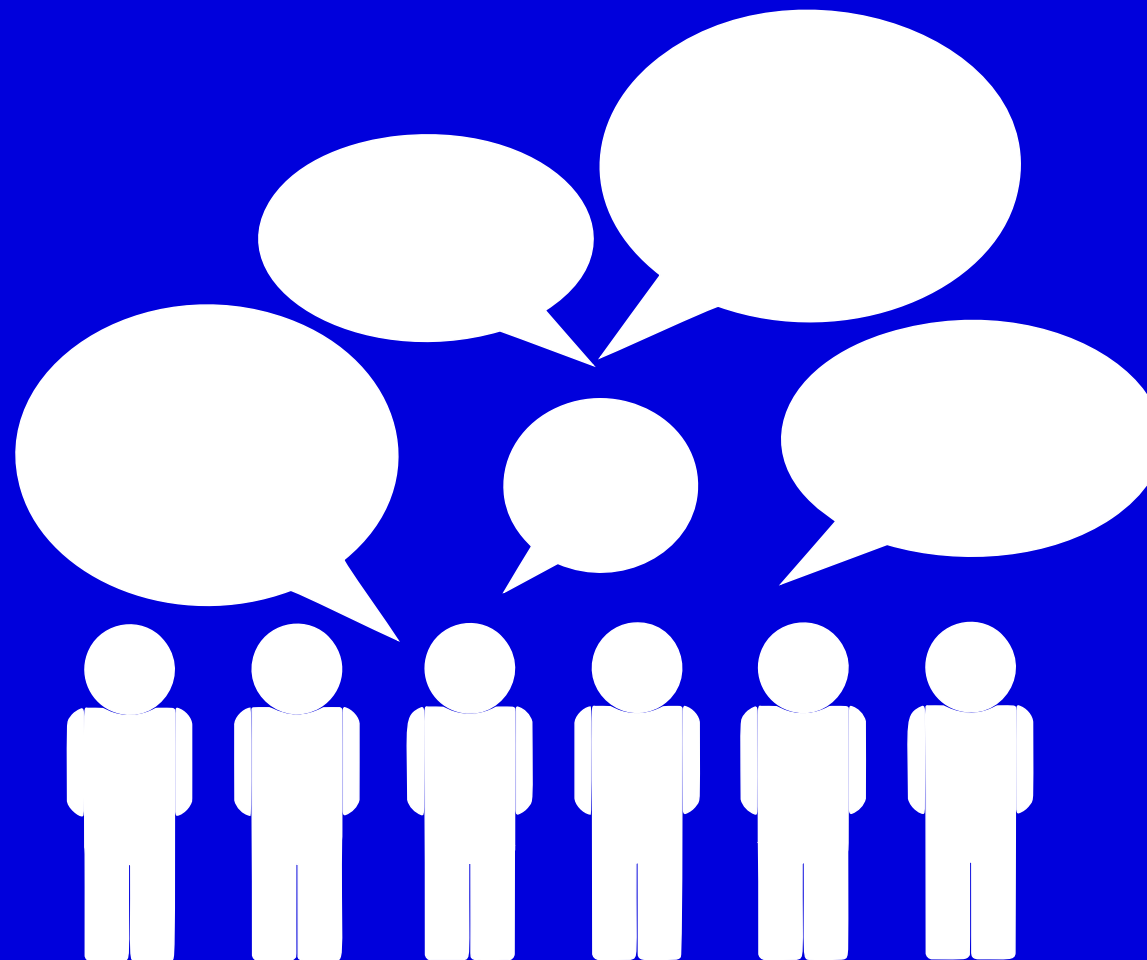
- Změna názvu na: *Data – odpověď na základní otázku života, vesmíru a vůbec...*
- Výuka **každý semestr**.
- **Čas** začátku a **místo** výuky?
- **Kanály pro propagaci předmětu** na fakultách?
- **Fyzicky, hybridně, nebo on-line?**
- Změnit **pořadí přednášek?**
- Posunout/zjednodušit/rozšířit **obsah** některé **přednášky?**

## – Vaše nápady:

- Říkejte.
- Pište na [core42@space.muni.cz](mailto:core42@space.muni.cz).

# MUNI

## Diskuse



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