

**MUNI**  
**ICS**

**FAIR data pro uživatele CF MAFIL**

# **Obecný úvod k FAIR datům, podpora na MUNI, EOSC-CZ**

Michal Růžička, ÚVT MU  
[ruzicka@ics.muni.cz](mailto:ruzicka@ics.muni.cz)

2024-02-14

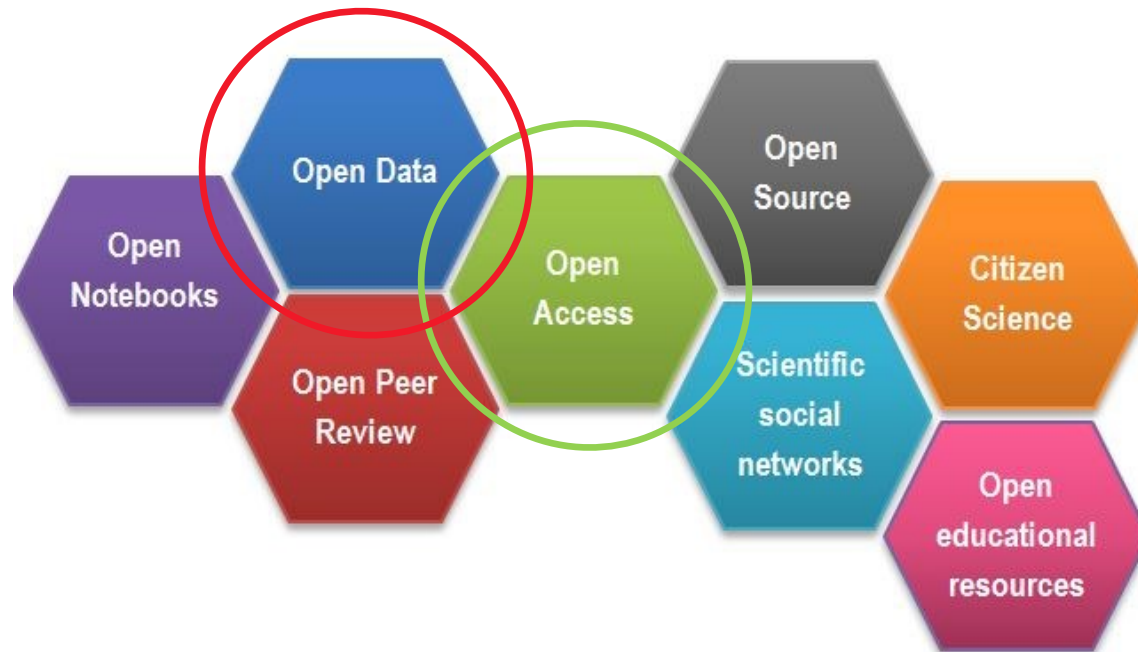
# Obsah přednášky

1. Kontext
2. Životní cyklus výzkumných dat
3. Úložiště pro výzkumná data v instituci a infrastruktura
  - Masarykova univerzita
  - CESNET
4. Repozitáře pro výzkumná data praktičtěji
  - Typy repositářů
  - Jak najít vhodný repositář
5. EOSC
  - Implementace EOSC v ČR
6. Kam dál
  - Kurz Bezpečná správa výzkumných dat
  - Kybernetická bezpečnost

**MUNI**  
ICS

# Kontext

# Schéma Open Science



Zdroj: <https://www.fosteropenscience.eu/node/1420>

### Politika Open Access

- **OA1:** Uchovávání vědeckých publikací MU
- **OA2:** Zvýšení dostupnosti vědeckých publikací MU

### Specializovaná podpora pro rozvoj Open Access

- **OA3:** Repozitář MU
- **OA4:** Vydávání otevřených publikací

### Financování Open Access

- **OA5:** Fond OA MU
- **OA6:** Adaptace EIZ na transformační model

### Politika FAIR Data

- **FD1:** Strategie pro správu a zpřístupňování výzkumných dat MU
- **FD2:** Evidence výzkumných dat

### Specializovaná podpora pro rozvoj FAIR Data

- **FD3:** Datový repozitář MU
- **FD4:** Podpora pro správu a zpřístupňování výzkumných dat

### Infrastruktura FAIR Data

- **FD5:** Infrastruktura pro výzkumná data

### Podpora pro vědce

- **OS1:** DMP+
- **OS2:** Lidské zdroje
- **OS3:** Vzdělávání a informovatnost
- **OS4:** Služby Open Science

### Interní procesy

- **OS5:** Politiky na jednotlivých HS
- **OS6:** Open Science v kontextu evaluace vědy
- **OS7:** Open Science a grantová schémata MU

### Standardy

- **OS8:** DOI
- **OS9:** ORCID
- **OS10:** OpenAIRE
- **OS11:** AOAP, ACAN

### Spolupráce

- **OS12:** Národní iniciativy
- **OS13:** Mezinárodní iniciativy

# Open Science a projekty

## – Podmínky:

- Horizon Europe
- Exceles
- OP JAK – Špičkový výzkum

## – Povinné/nepovinné principy

- **Povinné**: Open Access, Data Management (DMP)
- **Nepovinné**: Citizen Science, Open Source, Preprints, ...

# Životní cyklus výzkumných dat

# Životní cyklus výzkumných dat

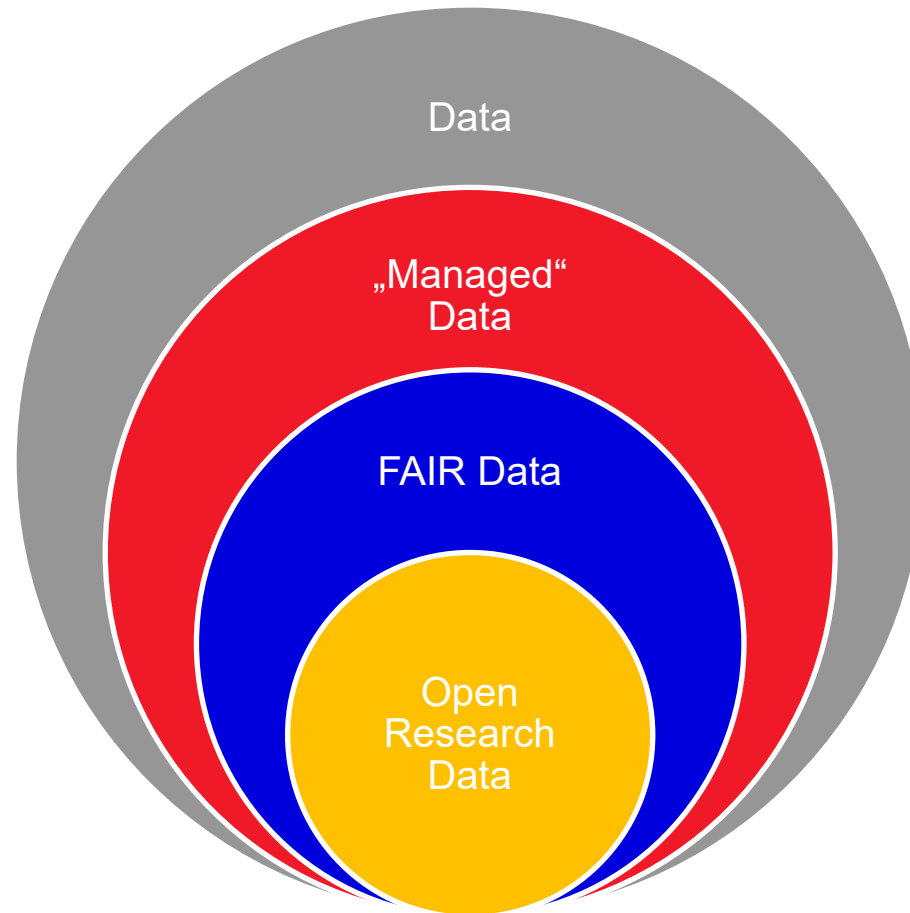


- Jaká data (znovu) používáte
  - včetně licencí, které vám to umožňují,
- jaká data generujete a jakým způsobem,
- kde je ukládáte, zálohujete, dlouhodobě uchováváte,
- jak je trvale a jedinečně identifikujete,
- jak je zpracováváte,
- jak je analyzujete,
- kde je zveřejňujete a sdílíte,
- kdo tohle 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.

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



# Úrovně dat



MUNI  
ICS

# Úložiště pro výzkumná data v instituci a české eInfrastruktuře

Masarykova univerzita, CESNET

# Kategorizace úložišť na MU a doporučení pro jejich využívání

- Datová úložiště na MU: <https://it.muni.cz/kategorie/datova-uloziste>
- Doporučení pro užívání úložišť: <https://it.muni.cz/prehledy/doporuzeni-pro-uzivani-ulozist>
  - Kategorizace dat.
  - Kategorizace úložišť.
  - Vhodnost úložišť pro různé typy dat.

# Úložiště na MU – Vhodnost úložišť pro různé typy dat

TYP ÚLOŽIŠTĚ	POUŽITÍ			
	ZELENÁ: VEŘEJNÁ DATA	MODRÁ: INTERNÍ DATA	ORANŽOVÁ: DISKRÉTNÍ DATA	ČERVENÁ: CITLIVÁ DATA
PŘENOSNÁ MÉDIA (FLASH DISKY, EXTERNÍ HDD, CD, DVD, ...)	Vhodné	Možné doporučeno použití šifrování	Nevhodné možné při použití šifrování	Nevhodné
LOKÁLNÍ ÚLOŽIŠTĚ				
V POČÍTAČÍCH (STOLNÍ, NOTEBOOKY)	Vhodné	Vhodné	Vhodné doporučeno použití šifrování	Nevhodné možné v dobře odůvodněných případech, při provedení individuální analýzy, použití šifrování a aplikaci dalších bezpečnostních opatření, která z analýzy vyplynou
V MOBILNÍCH ZAŘÍZENÍCH (MOBILNÍ TELEFONY, TABLETY, ...)	Vhodné	Vhodné nutný zámek obrazovky (vzor, čtečka otisků prstů, PIN, heslo)	Možné nutné použití šifrování nutný silný zámek obrazovky (čtečka otisků prstů, PIN, heslo)	Nevhodné možné v dobře odůvodněných případech, při provedení individuální analýzy, použití šifrování a aplikaci dalších bezpečnostních opatření, která z analýzy vyplynou
SÍŤOVÁ A CLOUDOVÁ ÚLOŽIŠTĚ ÚVT (TZV. STANDARDNÍ A STŘEDNÍ SÍŤOVÉ ÚLOŽITĚ, VIZ <a href="#">KATALOG IT, ÚLOŽIŠTĚ CERIT-SC</a> )	Vhodné	Vhodné	Vhodné	Vhodné, doporučeno provedení individuální analýzy, použití šifrování a aplikaci dalších bezpečnostních opatření, která z analýzy vyplynou

# Úložiště na MU – Vhodnost úložišť pro různé typy dat (2)

TYP ÚLOŽIŠTĚ	POUŽITÍ			
	ZELENÁ: VEŘEJNÁ DATA	MODRÁ: INTERNÍ DATA	ORANŽOVÁ: DISKRÉTNÍ DATA	ČERVENÁ: CITLIVÁ DATA
ÚLOŽIŠTĚ IS MUNI (NAPŘ. DOKUMENTOVÝ SERVER, ÚSCHOVNA APOD.)	Vhodné	Vhodné	Vhodné	Vhodné, doporučeno provedení individuální analýzy, použití šifrování a aplikaci dalších bezpečnostních opatření, která z analýzy vyplynou
ÚLOŽIŠTĚ CESNET (NAPŘ. CESNET ARCHIVNÍ ÚLOŽIŠTĚ, OWNCLOUD, FILESENDER, ..., VIZ <a href="#">ODDĚLENÍ DATOVÝCH ÚLOŽIŠŤ CESNET</a> )	Vhodné	Vhodné	Vhodné	Vhodné, doporučeno provedení individuální analýzy, použití šifrování a aplikaci dalších bezpečnostních opatření, která z analýzy vyplynou
EXTERNÍ ÚLOŽIŠTĚ				
SE SMLOUVOU S MUNI				
MUNI MICROSOFT O365 (MUNI O365 ONEDRIVE, SHAREPOINT, ..., VIZ <a href="#">MUNI O365</a> )	Vhodné	Vhodné	Vhodné doporučeno použití šifrování	Možné  výhradně s adekvátním procesním pokrytí dané situace na základě individuální analýzy a aplikaci bezpečnostních opatření, která z analýzy vyplynou
MUNI GOOGLE G SUITE FOR EDUCATION (VIZ <a href="#">MUNI GOOGLE APPS</a> )	Vhodné	Vhodné	Nevhodné možné při použití šifrování	Nevhodné
GRAMMARLY	Vhodné	Vhodné	Nevhodné	Nevhodné
BEZ SMLOUVY S MUNI				
VEŘEJNÁ GOOGLE, MICROSOFT, DROPBOX, ÚLOŽIŠTĚ	Vhodné	Nevhodné	Nevhodné	Nevhodné

# SensitiveCloud – infrastruktura pro práci s citlivými daty

- Budováno CERIT-SC.

- <https://www.cerit-sc.cz/infrastructure-services/sensitivecloud>

- Dvě komponenty:

- SensitiveCloud **Computer**.

- Poskytovaný jako *Platform as a Service (PaaS)*, postaveno na **Kubernetes**.

- Uživatel spravuje **jen aplikaci**, nikoliv celý virtuální stroj.

- Ideálně *Software as a Service (SaaS)* – R Studio, Jupyter Notebook apod.

- SensitiveCloud **Storage**.

- Úložiště pro zpracovávaná **data**.

- Integrace v rámci compute.

- Oddělená síť, WireGuard VPN, Perun AAI s vícefaktorovým přihlášením, fyzické zabezpečení hardware, ...

- **Formalizované** postupy správy, neustálé **vylepšování**, budováno dle požadavků **ISO 27k**.

# Oddělení datových úložišť CESNET

- <https://du.cesnet.cz/>
- Mohou být využívána pro účely výzkumu, a to akademickými pracovníky a studenty výzkumných institucí v ČR.
- Datová úložiště jsou certifikována podle normy pro systém managementu bezpečnosti informací ČSN EN ISO/IEC 27001:2014.
- Případy použití:
  - Datový prostor pro zálohování.
  - Datový prostor pro archivaci.
  - Datový prostor pro sdílení dat.

MUNI  
ICS

# Repozitáře pro výzkumná data

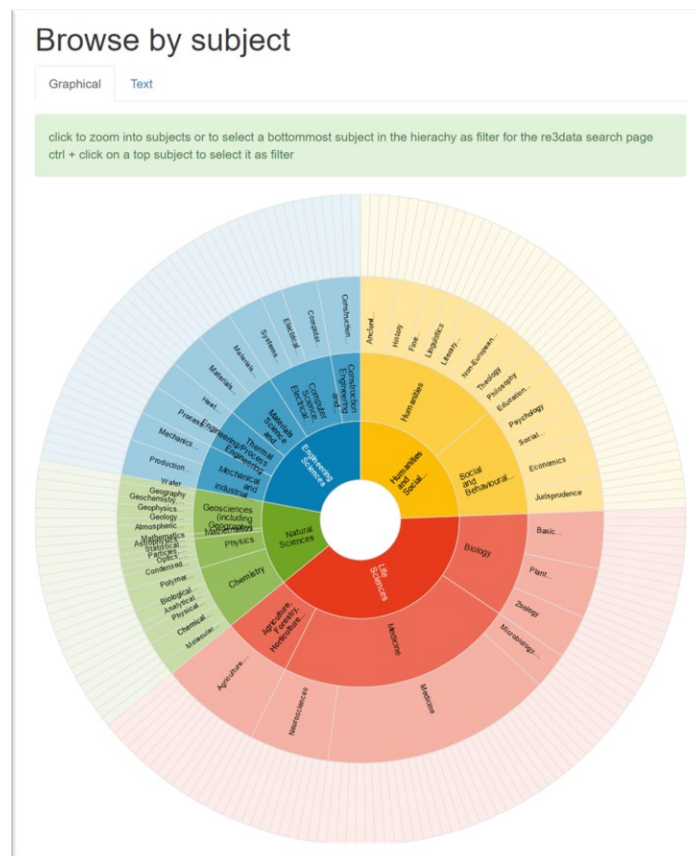


# Typy repozitářů

- **Oborové, národní, institucionální, catch-the-all, ...**
- Běžný postup práce s výzkumnými daty je **dělení po oborech, nikoliv po státech či institucích.**
  - Nezohledňujeme národní/institucionální/... specifika, ale oborová specifika.
- Pokud řešíme projektový/institucionální/národní **standard**, musíme být **kompatibilní s mezinárodními oborovými zvyklostmi.**
  - Oborové standardy řeší např. EOSC Task Forces: <https://www.eosc.eu/task-force-faq>
- Pokud chceme shromažďovat **informace napříč obory**, musí to být **formát dostatečně obecný**, který možná přesáhne rámec výzkumných dat.
  - Např. aplikační profil formátu DCAT ([DCAT-AP](#)), který se používá pro popis otevřených dat např. ze státní správy.

# Jak najít/vybrat datový repozitář

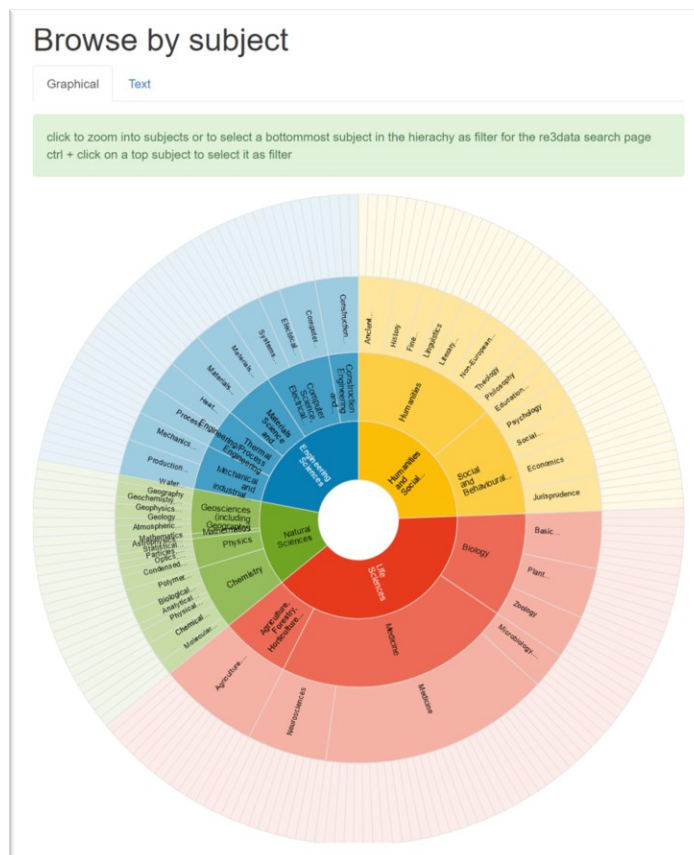
- OpenAIRE: [Jak najít důvěryhodný repozitář pro vaše data](#)
  - Preferovány jsou důvěryhodné certifikované repozitáře.
    - [CoreTrustSeal](#) (seznam certifikovaná repozitářů).
    - [Nestor Seal](#) (verifikace dle DIN 31644).
    - [ISO 16363](#).
  - Např. ale známé [Zenodo](#) žádnou certifikaci nemá...
- Mezi nejpoužívanější obecné repozitáře patří
  - [Zenodo](#),
  - [Figshare](#) nebo
  - [Dryad](#).



Zdroj: <https://www.re3data.org/browse/by-subject/>

# Jak najít/vybrat datový repozitář (2)

- Adresáře repozitářů:
  - Open Access repozitáře: [OpenDOAR](https://open.doar.org/)
  - Datové repozitáře: [re3data.org](https://re3data.org/)
- [ELIXIR RDMkit | Data publication](https://elixir-europe.org/ELIXIR-RDMkit/Data-publication)
  - [EMBL-EBI's data submission wizard](https://www.ebi.ac.uk/submit/).
  - [ELIXIR Deposition Databases](https://www.ebi.ac.uk/submit/).
  - [Scientific Data journal's recommended repositories](https://www.sciencedirect.com/journal/scientific-data).



Zdroj: <https://www.re3data.org/browse/by-subject/>

# Na co dát pozor při výběru repozitáře

- Přidělí repozitář vašim datům **perzistentní identifikátor** (např. DOI)?  
Díky perzistentním identifikátorům jsou vaše data snáze vyhledatelná a citovatelná.
- Je repozitář **důvěryhodný/certifikovaný**?  
U certifikovaných repozitářů máte větší jistotu, že je o vaše data dobře postaráno.
- Poskytuje vybraný repozitář **otevřený přístup** k uloženým datům?  
Pokud chcete svá data sdílet otevřeně, pak je toto klíčová informace.
- Opatří repozitář vaše data **licencí**? Uvádí jasné podmínky, za jakých lze data uložená v repozitáři využívat?  
Je důležité, aby uživatelé vašich dat věděli, jak s nimi mohou nakládat.
- Poskytne repozitář pro vaše data **úvodní stránku s metadaty** (tzv. landing page)?  
Metadata pomohou ostatním data najít, zjistit, co je to za data, a také jak je citovat.
- Umožňuje repozitář **verzování**?  
Pokud svůj dataset aktualizujete, můžete aktualizovanou verzi nahrát jako novou verzi k původnímu datasetu. Nový dataset dostane přidělený svůj vlastní perzistentní identifikátor a uživatelé snadno zjistí, jaká je nejnovější verze nebo jaká verze byla použita v původní studii.

**MUNI**  
ICS

**EOSC**

# Implementace European Open Science Cloud (EOSC) v ČR

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

National Czech Programme | **eosc**

Hledej...  English

O nás Pracovní skupiny Sekretariát EOSC-CZ Školící centrum Novinky a akce Kontakt

## EOSC-CZ

European Open Science Cloud (EOSC) je evropská iniciativa zaměřená na rozvoj infrastruktury podporující postupy otevřené vědy v oblasti správy výzkumných dat. Nyní se začíná implementovat v ČR.

# EOSC v ČR – očekávané služby Národní datové infrastruktury (NDI)

- Služby očekávané od NDI vychází z výstupů pracovní skupiny Základní služby.
  - Pracovní skupiny pro implementaci EOSC v ČR jsou stále otevřené novým členům, viz Pracovní skupiny implementace EOSC v ČR.
- Navrhované služby se dají rozdělit do několika základních kategorií:
  - Důvěryhodné repozitáře
  - Služby pro podporu práce s persistentními identifikátory (PID)
  - Služby pro řízení přístupu k datům, autentizaci a autorizaci uživatelů (AAI) a podporu jejich spolupráce
  - Služby pro podporu a plánování správy výzkumných dat
  - Machine actionability
  - Služby pro FAIRifikaci výzkumných dat
  - Služby transferu dat
  - Kvalita služeb
  - Kompatibilita s existujícími platformami
  - Statistiky a přehledy
  - Národní příspěvek k budování EOSC

# Důvěryhodné repozitáře

- Národní repozitářová platforma (NRP) nad NDI.
  - Součástí i Národní metadatový adresář (NMA).
- Místo, kam můžete (i velká) data nejen uložit, ale také
  - spolehnout se, že tam budou v následujících letech dostupná a
  - budete moci s daty pracovat.
- Určeno i pro „živá“ data, tj. podpora aktivní práce s daty, ne jen uložení archivních dat.
- Poskytování služeb dle vhodných standardů/certifikací pro provoz důvěryhodných repozitářů.
  - Důvěryhodné repozitáře dle definic EU, EOSC, grantových agentur apod.
  - Pravděpodobně CoreTrustSeal.
  - Rodina standardů ISO 27 tisíc apod.?
- Nestaráte se o vlastní infrastrukturu, zabezpečená úložiště apod., ale
  - využíváte služby.
  - NDI umožní postavení vlastních instancí repozitářů (oborových, institucionálních, ...).



# Precizní kontrola vlastníka dat nad řízením přístupu k datům i metadatům

- Data uložená v NDI jsou pod plnou kontrolou jejich vlastníků.
  - Zpřístupnění dalším osobám se děje na pokyn jejich vlastníka.
- Už samotná informace o existenci dat či metadata může být citlivá informace.
  - Řízení přístupu až na úrovni řízení viditelnosti existence samotných metadat či jejich částí apod.
- Federované AAI služby
  - Autentizace uživatelů přes federované služby přihlašování (eduID.cz apod.),
  - tj. s využitím existující identity z instituce, bez nutnosti vytváření dalšího účtu, hesla atd.

# Služby pro sdílení dat s jinými uživateli

- Sdílení dat by mělo být možné jednak s konkrétními osobami a skupinami
  - výběr např. z institucionálních identit daných federovaným přihlášením apod.,
  - případně sdílením „odkazem“.
- Může být řešeno i formou „data visiting“.
  - Řízené povolením přístupu vybraným uživatelů do „domovského“ úložiště s danými daty,
  - bez nutnosti přenosu do jiné lokality.
- Možnost anonymních view-only přístupů k datům.
  - Pro podporu předání dat pro zaslepené recenzní řízení apod.
  - Stále častější povinnost sdílení podkladových dat k publikacím.
    - Někdy již v okamžiku recenzního řízení.
    - Ukázka existující implementace např. v platformě OSF: <https://help.osf.io/hc/en-us/articles/360042097853-Create-a-View-only-Link-for-a-Registration>

# Služby pro podporu a plánování správy výzkumných dat

- Definované (strojově zpracovatelné) politiky práce s daty v repozitářích v NDI.
- Podpora tvorby DMP.
  - Vygenerování DMP
    - nebo vybraných částí
  - na základě politik použitého repozitáře, znalosti licencí, formátů, velikostí dat, cílených retencí dat (maximální či minimální doba uchovávání např. na základě zákonných požadavků) apod.
  - Možnost napojení/integrace existujících nástrojů pro DMP.
    - Např. Data Stewardship Wizard.

# Metriky a měření plnění FAIR atributů pro výzkumná data

- Pro datasey může být zajímavé nějakým způsobem **měřit** plnění jednotlivých **FAIR atributů**.
  - NRP EOSC CZ by měla **definovat metriky** hodnocení a **poskytovat nástroje** pro **měření** plnění FAIR atributů.
- Hodnocení na základě informací o
  - datech,
  - metadatech,
  - nastavení a politikách konkrétního repozitáře apod.
- Příkladem implementace podobného nástroje může být např. FAIR-Checker (<https://fair-checker.france-bioinformatique.fr/>) nebo FAIR metriky pro DMP v Data Stewardship Wizard.



Zdroj: <https://researchers.ds-wizard.org/>

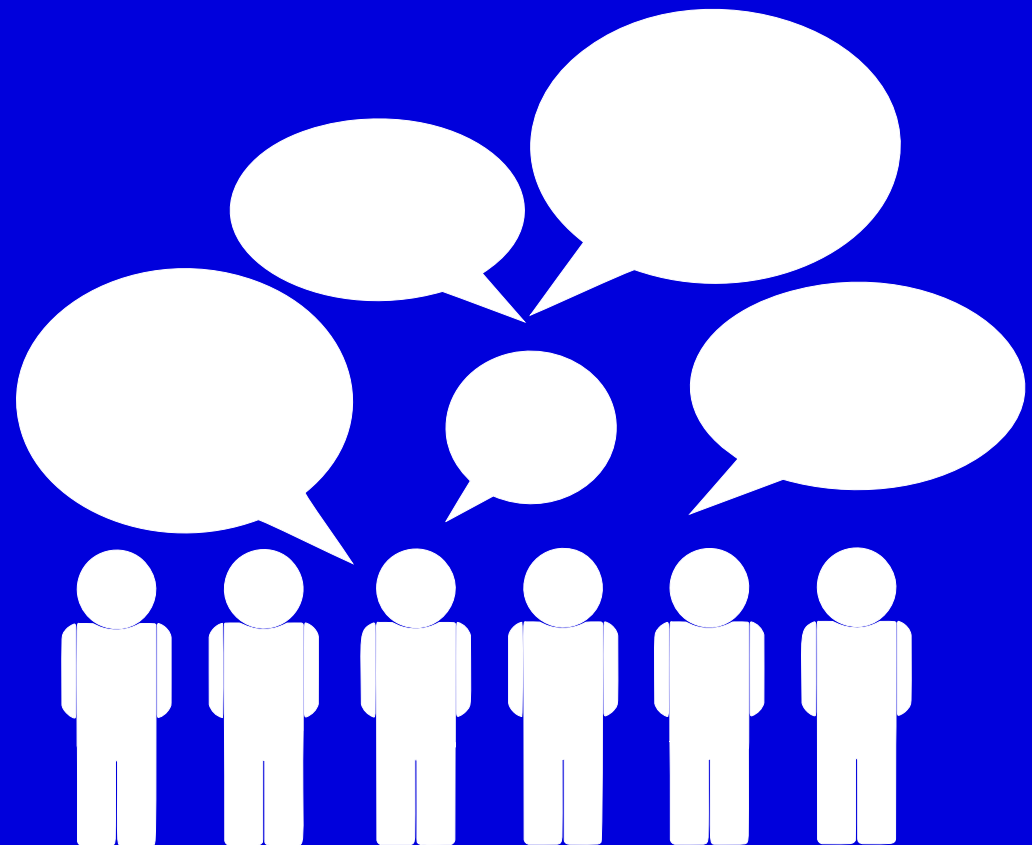
**MUNI**  
ICS

**Kam dál**

# Kam dál – další zdroje informací

- Kyberbezpečnostní doporučení týmu CSIRT-MU
  - <https://csirt.muni.cz/>
  - Bezpečnostní portál MUNI
    - <https://security.muni.cz/>
  - **Kurz Kyberkompas**
    - <https://security.muni.cz/cybercompass>
  - Kurz GDPR
    - <https://security.muni.cz/gdpr-kurz>
  - Techniky sociálního inženýrství
    - [https://security.muni.cz/socialni\\_inzenyrstvi](https://security.muni.cz/socialni_inzenyrstvi)
- Pro vaše studenty: Kurz *CORE042 Data – odpověď na základní otázku života, vesmíru a vůbec...*
  - <https://is.muni.cz/predmet/cus/jaro2024/CORE042>
  - Běh Jaro 2024 **startuje** úvodní přednášku už v úterý 20. 2. 2024!

# Otázky?



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