

# Přístup vědců ke správě výzkumných dat na Mendelově univerzitě v Brně

## Výsledky průzkumu

Zpracovali: Mgr. Naděžda Fasurová, Ph.D., Mgr. Martina Šmardová,

Ústav vědecko-pedagogických informací a služeb, Mendelova univerzita v Brně

V březnu 2022 proběhl na univerzitě průzkum “Přístup vědců ke správě výzkumných dat na MENDELU”. Cílem bylo zmapovat, jak vědci zacházejí s daty a jaké mají problémy a potřeby v této oblasti. Dotazník byl prvním krokem, jak analyzovat situaci na univerzitě.

Byl použit volně dostupný dotazník univerzity VŠB-TUO, cílovou skupinou byli akademičtí a vědečtí pracovníci univerzita a studenti doktorského studia.

**Z odpovědí respondentů a ze zhodnocení dotazníku vyplynuly tyto výsledky a doporučení:**

- Respondenti požadují nastavení správy výzkumných dat na univerzitě, vznik doporučení pro výzkumná data na univerzitě.
- Očekávají, že univerzita nabídne možnosti pro ukládání a sdílení (živých) dat během výzkumu (pro různé formáty a objemy, možnost sdílení i mimo univerzitu).
- a že univerzita zajistí bezpečné a trvalé uchovávání dat po skončení projektů, včetně metadat a identifikátoru DOI.

Organizátoři dotazníku doporučují řešit výzkumná data komplexně napříč univerzitními pracovišti, a to ve spolupráci Odboru pro podporu tvůrčí činnosti, Odboru informačních technologií, Výzkumného centra, Ústavu vědecko-pedagogických informací a služeb. Nepominutelný by měl být názor a zkušenosti projektových manažerů HORIZON projektů.

**Mezi potřebné činnosti patří:**

1. zapojit univerzitu do aktivit EOSC (<https://www.e-infra.cz/eosc/pracovni-skupiny-pro-implementaci-eosc-v-cr>) a do národní datové infrastruktury (projekt EOSC <https://www.cesnet.cz/vyzkum-a-vyvoj/eosc/>).
2. poskytovat IT technickou podporu pro práci s výzkumnými daty a podporu pro práci s nástroji a infrastrukturou pro ukládání dat (úložiště a sdílení dat), vytvořit doporučení pro používání uložišť
3. využít experty z jiných organizací, kteří mají zkušenosti se správou dat a čerpat inspiraci u jiných výzkumných organizací
4. vytvořit Open Science centrum a vytvořit potřebné pracovní pozice metodik správy výzkumných dat (Data Steward) a garant pro nastavení infrastruktury ukládání dat. Cílem je nabízet metodickou podporu a vzdělávání v oblasti správy výzkumných dat.

## ÚVOD DO PROBLEMATIKY Open Science

[Open Science](#) je otevřený přístup k vědeckým informacím v širším slova smyslu, zahrnuje více témat: otevřený přístup k publikacím (Open Access), otevřená data (Open Data), otevřené zdrojové kódy (Open Source), otevřené recenzní řízení (Open Peer Review), otevřené vzdělávací materiály (Open Education), alternativní metriky hodnocení (Alternative Metrics), zapojení občanů do vědy (Citizen Science) a další. Otevřená věda nabízí větší transparentnost a možnost replikovat vědu, vědecké poznání je rychle dostupné široké veřejnosti.<sup>1</sup>

Open Access znamená otevřený přístup k vědeckým informacím. Mendelova univerzita v Brně ve svém dlouhodobém záměru pro rok 2021 a dále podporuje otevřený přístup (prioritní cíl 4, s. 24), ale také nastavení procesů pro Open Science<sup>2</sup>.

Z Národní strategie otevřeného přístupu a z cílů Akčního plánu se řeší tyto body:<sup>3</sup>

- podpora zlaté cesty otevřeného přístupu k vědeckým informacím úhradou APC (Article Processing Charge) poplatků,
- podpora zelené cesty otevřeného přístupu k vědeckým informacím rozvojem univerzitního otevřeného repozitáře výsledků vědy a výzkumu,
- zahájení změny vědeckého prostředí tak, aby se otevřenost přístupu k vědeckým poznatkům stala postupně běžnou praxí vědeckých pracovníků,
- Vytvoření strategie Open Science na univerzitě.

## Správa výzkumných dat

Správa výzkumných dat ([Research Data Management](#)<sup>4</sup>, RDM) popisuje způsob, jak efektivně organizovat, ukládat a uchovávat data, která byla získána během výzkumného projektu a stává se nedílnou součástí vědecké praxe. Životní cyklus vědeckých dat:



Obrázek č. 1: ELIXIR-CONVERGE, RDM Kit, zdroj obrázku: <https://rdmkit.elixir-europe.org>, obrázek je dostupný pod licencí Creative Commons Attribution 4.0 International License.

<sup>1</sup> <https://vedavyzkum.cz/nazory-a-komentare/nazory-a-komentare/eva-hnatkova-open-science-zbozne-prani-nebo-realna-budoucnost-vedy>

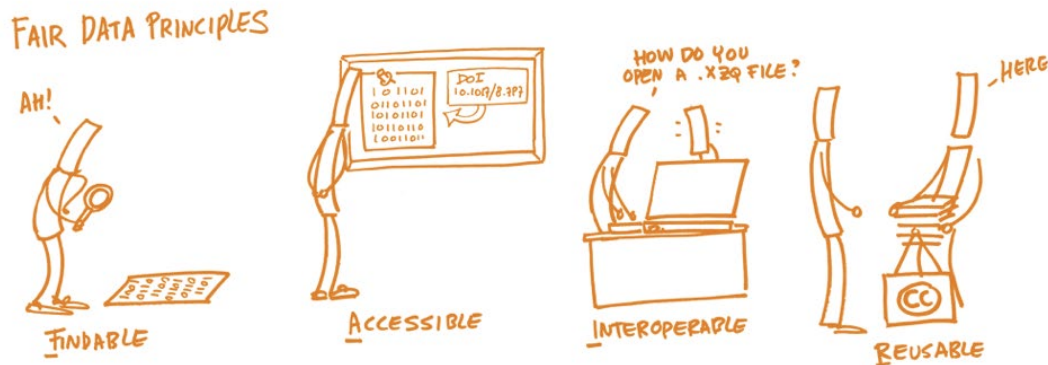
<sup>2</sup> <https://mendelu.cz/o-univerzite/uredni-deska/uredni-deska-dlouhodoby-zamer-univerzity/?psn=0>

<sup>3</sup> <https://www.vyzkum.cz/FrontClanek.aspx?idsekce=876326>

<sup>4</sup> <https://rdmkit.elixir-europe.org/>

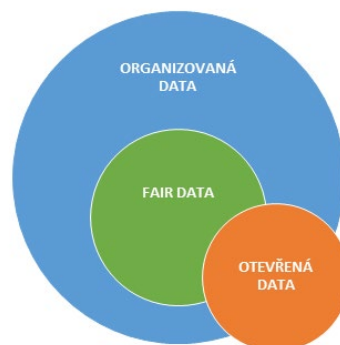
## FAIR data

Výzkumná data by měla být [FAIR](#)<sup>5</sup>, tedy spravována v souladu s [principy FAIR](#)<sup>6</sup>: Findable-dohledatelná, Accessible-dostupná pro lidi i strojově čitelná, Interoperable-interoperabilní Reusable-znovu využitelná.



Obrázek č. 2: The Open Science Training Handbook, zdroj obrázku: <https://doi.org/10.5281/zenodo.1212496>, obrázek je licencován pod licencí [CC0 1.0 Universal](#).

FAIR data neznamenají automaticky otevřená data. Výzkumná data lze otevřít v těch případech, kdy je to možné: **“as open as possible, as closed as necessary”**. Překážkou pro plné otevření dat nejčastěji jsou citlivá data a osobní údaje (GDPR), smlouvy s třetí stranou (komerční výzkum) nebo podání patentové žádosti.



Obrázek č. 3: FAIR data a Otevřená data., zdroj obrázku: vlastní.

## Otevřená data

Články v časopisech vznikají na základě výzkumné činnosti, naměřených anebo zjištěných dat. Sdílení výzkumných dat umožňuje ověřit tvrzení a závěry ve vědeckých člancích (reprodukovatelnost vědy) anebo tato data znovu využít v dalším výzkumu. V současné době jak v evropském prostředí, tak i na národní úrovni ze strany donátorů financí se zesiluje tlak na zveřejňování výzkumných dat financovaných z veřejných zdrojů v úložištích, v datových repozitářích a také v časopisech, které zveřejňují výzkumná data.

<sup>5</sup> <https://www.go-fair.org/fair-principles/>

<sup>6</sup> <https://www.openaire.eu/how-to-make-your-data-fair>

## Správa výzkumných dat na univerzitách v České republice

Univerzita Karlova v rámci Centra na podporu Open Science poskytuje **metodickou podporu k Open Science** a k managementu výzkumných dat, nabízí také e-learningový kurz [Správa výzkumných dat](#)<sup>7</sup>. Na Masarykově univerzitě bylo zřízeno metodické pracoviště [Open science centrum](#)<sup>8</sup>, které kromě jiného poskytuje také metodickou podporu k výzkumným datům. V roce 2022 proběhl na Masarykově univerzitě kurz [Bezpečná správa výzkumných dat](#)<sup>9</sup>. V současné době se na některých českých vysokých školách připravují politiky pro výzkumná data nebo doporučení, jak s výzkumnými daty zacházet.

Česká republika je zapojena do **European Open Science Cloud (EOSC)**<sup>10</sup>. EOSC je nástroj k dosažení přístupnosti, interoperability a opětovné využitelnosti výzkumných dat a nabízí volně dostupné služby pro jejich uchování, správu, sdílení, analýzu a opakované použití pro všechny vědecké disciplíny. V rámci EOSC se připravuje Národní datová infrastruktura, vznikne národní metadatový adresář, národní repozitářová platforma a s tím spojené služby a vzdělávání. V rámci České republiky vznikl pilotní [národní datový repozitář](#)<sup>11</sup>, který nabízí vědcům možnost ukládat datové sady.

### Současný stav správy výzkumných dat na MENDELU

Na univerzitě zatím nejsou pravidla pro správu výzkumných dat ani doporučení pro akademické a vědecké pracovníky, jak mají s výzkumnými daty nakládat. Na jednotlivých ústavech pracují s výzkumnými daty dle zvyklostí ústavu a podle oboru výzkumu.

#### Datová infrastruktura na univerzitě

Běžnou praxí je ukládání dat na fyzických úložištích (počítač, externí disk, flash disk...), na síťovém disku nebo v cloudu. V případě fyzických úložišť bývají data uložena na více místech zároveň. U všech typů úložišť je preferováno šifrování dat a zaheslované přístupy silnými hesly delšími než 15 znaků.

Na univerzitě existuje možnost využít osobních stránek (user.mendelu.cz)<sup>12</sup>, připojení k serveru Disk Mendelu<sup>13</sup> a možnost založení projektových stránek s datovým úložištěm<sup>14</sup>. Požadavky na datová úložiště jsou na univerzitě řešeny individuálně. Odbor informační infrastruktury nabízí nastavení sdílených prostor výzkumným týmům a odborných konzultací při pořizování externích úložišť a odborné konzultace a pomoc při hledání konkrétních řešení už při tvorbě projektové žádosti.

---

<sup>7</sup> <https://openscience.cuni.cz/OSCI-112.html>

<sup>8</sup> <https://openscience.muni.cz/>

<sup>9</sup> <https://www.ics.muni.cz/o-nas/celozivotni-vzdelavani>

<sup>10</sup> <https://www.e-infra.cz/eosc>

<sup>11</sup> <https://data.narodni-repozitar.cz/>

<sup>12</sup> <https://tech.mendelu.cz/24978-osobni-stranky-uzivatelu>

<sup>13</sup> <https://tech.mendelu.cz/24980-pripojeni-k-serveru-disk>

<sup>14</sup> <https://tech.mendelu.cz/25024-pripojeni-projektovych-stranek>

Výzkumníci mohou využívat cloudové úložiště OneDrive<sup>15</sup> v rámci služby o365 od Microsoft, velikost je pro jednoho uživatele omezená na 1TB. Pro sdílení výzkumných dat lze požádat o zřízení týmového webu na platformě SharePoint<sup>16</sup>.

CESNET obecně nabízí všem výzkumníkům pro ukládání dat ownCloud<sup>17</sup> (omezeno na 100 GB na uživatele), ale tento způsob uložení není doporučeno používat pro důležitá data projektu<sup>18</sup>. Služba FileSender<sup>19</sup> pro sdílení dat nabízí dočasné uložení (v maximální velikosti 500 GB). Pro využívání těchto služeb od CESNET jsou potřeba pouze univerzitní přihlašovací údaje. Výzkumníci mohou také využít různé externí datové repozitáře (Zenodo apod.).

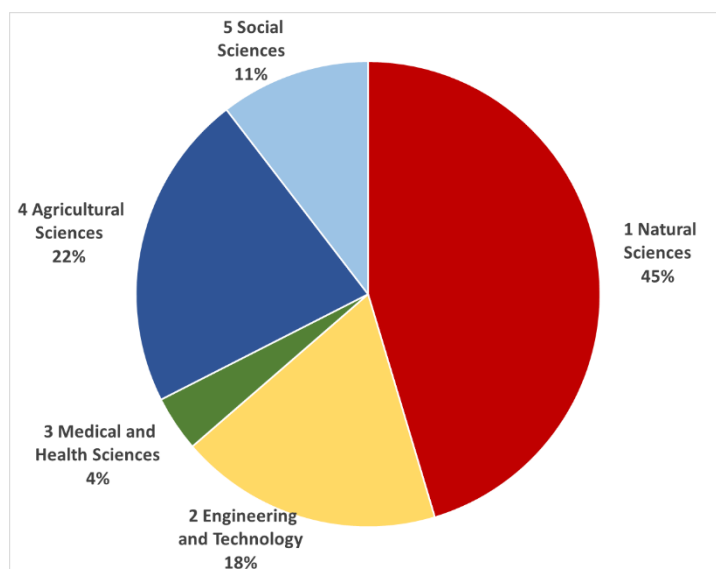
## Věda a výzkum na Mendelově univerzitě v Brně

Univerzita má 9 019 studentů, z toho 524 studentů doktorandského studia (FTE 2021). Na univerzitě působí 556 akademických pracovníků a 199 vědeckých a odborných pracovníků.

### Oborové rozdělení univerzity v databázích Web of Science a Scopus

Na základě analýzy oborového rozdělení výsledků univerzity z databáze Web of Science bylo zjištěno, že nejvíce výsledků je v kategoriích 1 Natural Sciences, 4 Agricultural Sciences a 2 Engineering and Technology. Nejvíce zastoupené obory u výsledků v databázi Scopus jsou stejné jako v databázi Web of Science, avšak liší se podílem v kategorii 2 Engineering and Technology.

1. Oborové rozdělení univerzity ve **Web of Science** podle počtu dokumentů (2020), zdroj: Web of Science, nástroj InCites. Kategorie 6 Arts and Humanities nebyla vyhodnocena kvůli nízkému počtu výsledků, podíl je pod 1 %.



Graf č. 1: Oborové rozdělení univerzity ve Web of Science za rok 2020

<sup>15</sup> <https://tech.mendelu.cz/30122-onedrive>

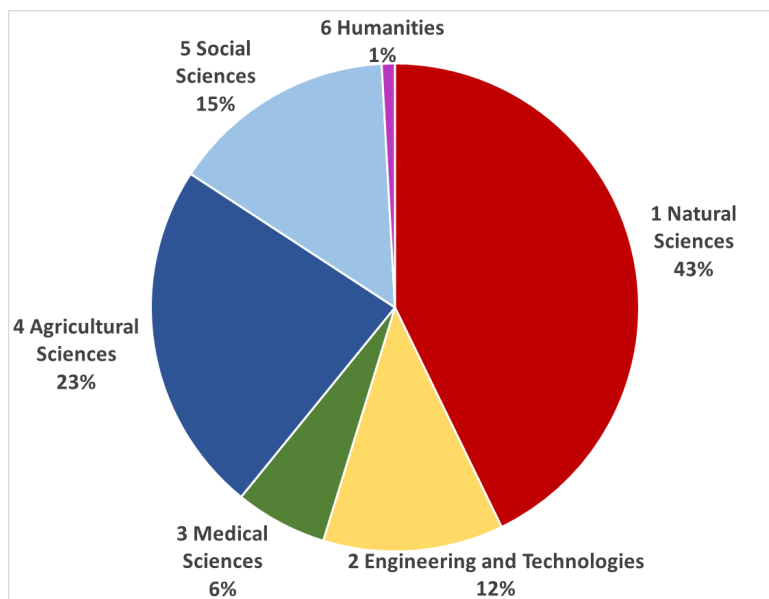
<sup>16</sup> <https://tech.mendelu.cz/25023-sharepoint>

<sup>17</sup> <https://tech.mendelu.cz/25052-owncloud>

<sup>18</sup> <https://it.ldf.mendelu.cz/?p=128>

<sup>19</sup> <https://tech.mendelu.cz/26361-filesender>

2. Oborové rozdělení univerzity ve **Scopus** podle počtu dokumentů (2020), zdroj: Scopus, nástroj SciVal.



Graf č. 2: Oborové rozdělení univerzity ve Scopus za rok 2020

Tabulka č. 1: Porovnání zastoupení oborů v citačních databázích Web of Science a Scopus

Oborová kategorie - Research Fields	Zastoupení oboru ve WoS (%)	Zastoupení oboru ve Scopus (%)
<b>1 Natural Sciences</b>	<b>45</b>	<b>43</b>
<b>2 Engineering and Technology</b>	<b>18</b>	<b>12</b>
3 Medical and Health Sciences	4	6
<b>4 Agricultural Sciences</b>	<b>22</b>	<b>23</b>
5 Social Sciences	11	15
6 Arts and Humanities	Méně než 1, nezapočteno	1

Publikace univerzity v oboru 5 Social Sciences jsou s podílem 15 % více zastoupeny v databázi Scopus než ve Web of Science, je třeba počítat i s výzkumnými daty tohoto oboru. Méně jsou na univerzitě zastoupeny obory 3 Medical and Health Sciences s podílem 5 % a obor 6 Arts Humanities, který tvoří přibližně 1 %. Dále vznikají na univerzitě výsledky v rámci mezioborového výzkumu.

## Dotazník: Přístup vědců ke správě výzkumných dat na MENDELU

V březnu 2022 proběhl na univerzitě průzkum “Přístup vědců ke správě výzkumných dat na MENDELU”. Cílem bylo zmapovat, jak vědci zacházejí s daty a jaké mají problémy a potřeby v této oblasti. Dotazník byl prvním krokem, jak analyzovat situaci na univerzitě. Byl použit volně dostupného dotazníku univerzity VŠB-TUO, který byl zpracován v Microsoft Forms. Podobný průzkum proběhl také na Univerzitě Karlově a Univerzitě Tomáše Bati.

### Metodologie a etika výzkumu

Dotazník organizoval Ústav vědecko-pedagogických informací a služeb (ÚVIS). Všem výzkumným pracovníkům a studentům doktorského studia byl odeslán e-mail s žádostí o vyplnění dotazníku, link (<https://uvis.mendelu.cz/article/dotaznikove-setreni-sprava-vyzkumnych-dat>) byl zveřejněn na webech ÚVIS. Dotazník obsahoval celkem 40 otázek, jednalo se jak o otázky uzavřené s možností výběru z více odpovědí, tak otevřené otázky, kde vědci mohli vyjádřit svůj názor. Dotazník byl anonymní. Organizátoři dotazníku se rozhodli neoslovovat studenty bakalářských a magisterských oborů. Důvodem bylo zaměření průzkumu přednostně na vědecké pracovníky, aby mohla vzniknout potřebná podpora pro správu výzkumných dat. Vyhodnocení proběhlo na ÚVIS.

### Šetření na MENDELU bylo zaměřeno na následující oblasti:

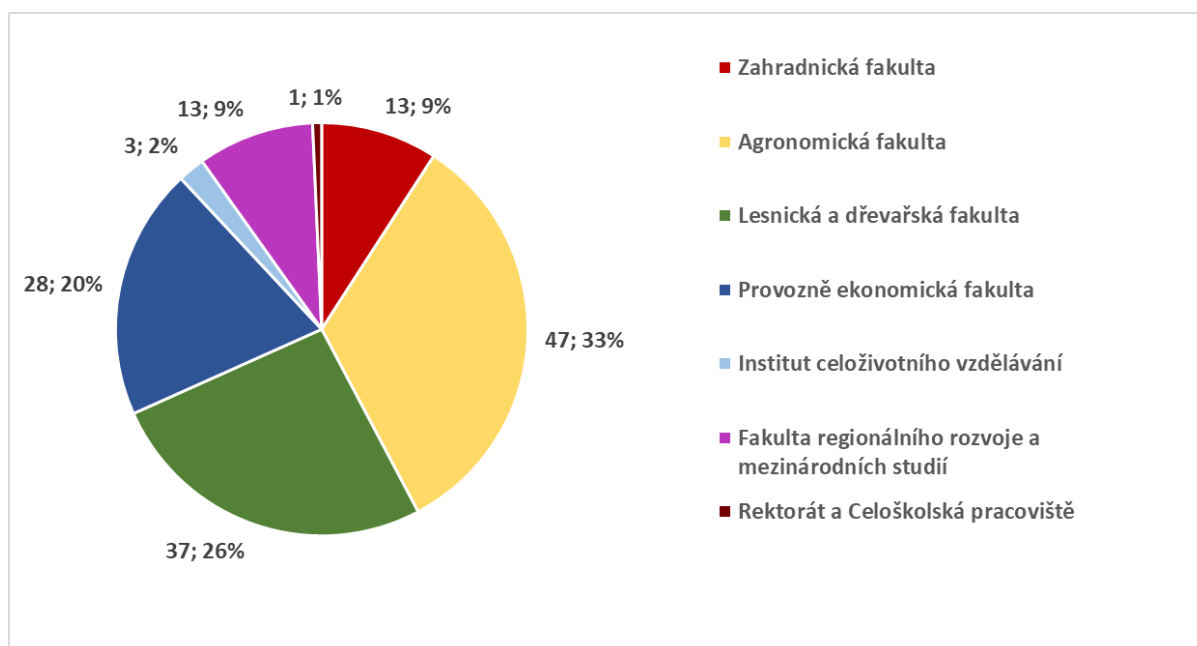
1. Údaje o respondentovi;
2. Práce s výzkumnými daty během výzkumu;
3. Dlouhodobé uchování výzkumných dat;
4. Sdílení výzkumných dat;
5. Podpora správy výzkumných dat.

### 1. Oblast dotazníku: Údaje o respondentovi

Nejprve byly položeny otázky zaměřené přímo na respondenty, tyto údaje slouží k identifikaci pracovního zařazení respondentů. Celkem **odpovědělo 142 respondentů** z celkového počtu 756 akademických a vědeckých pracovníků a 524 studentů doktorského studia na MENDELU (celkem 1280 FTE v roce 2021). FTE= Full Time Equivalent, ekvivalent plného pracovního úvazku.

#### Organizační jednotka

Největší odezva na dotazníkové šetření byla z Agronomické fakulty (47/142; 33,10 % všech respondentů), dále z Lesnické a dřevařské fakulty (37/142; 26,06 %) a z Provozně ekonomické fakulty (28/142; 19,72 %).



Graf č. 3: Počet a podíl respondentů z jednotlivých fakult.

Počet respondentů z fakult lze porovnat s aktuálními počty akademických, vědeckých pracovníků a doktorandů na jednotlivých fakultách. Z porovnání vyplývá, že nejvyšší % zastoupení respondentů vzhledem k celkovému počtu akademických, vědeckých pracovníků a doktorandů je z ICV, což ale není primární cílová skupina. Dále bylo zjištěno, že se průzkumu zúčastnilo poměrně velké procento respondentů z celkového počtu akademických, vědeckých pracovníků a doktorandů na FRRMS (24,53 %). U ostatních fakult se účast na průzkumu pohybuje okolo 10 %.

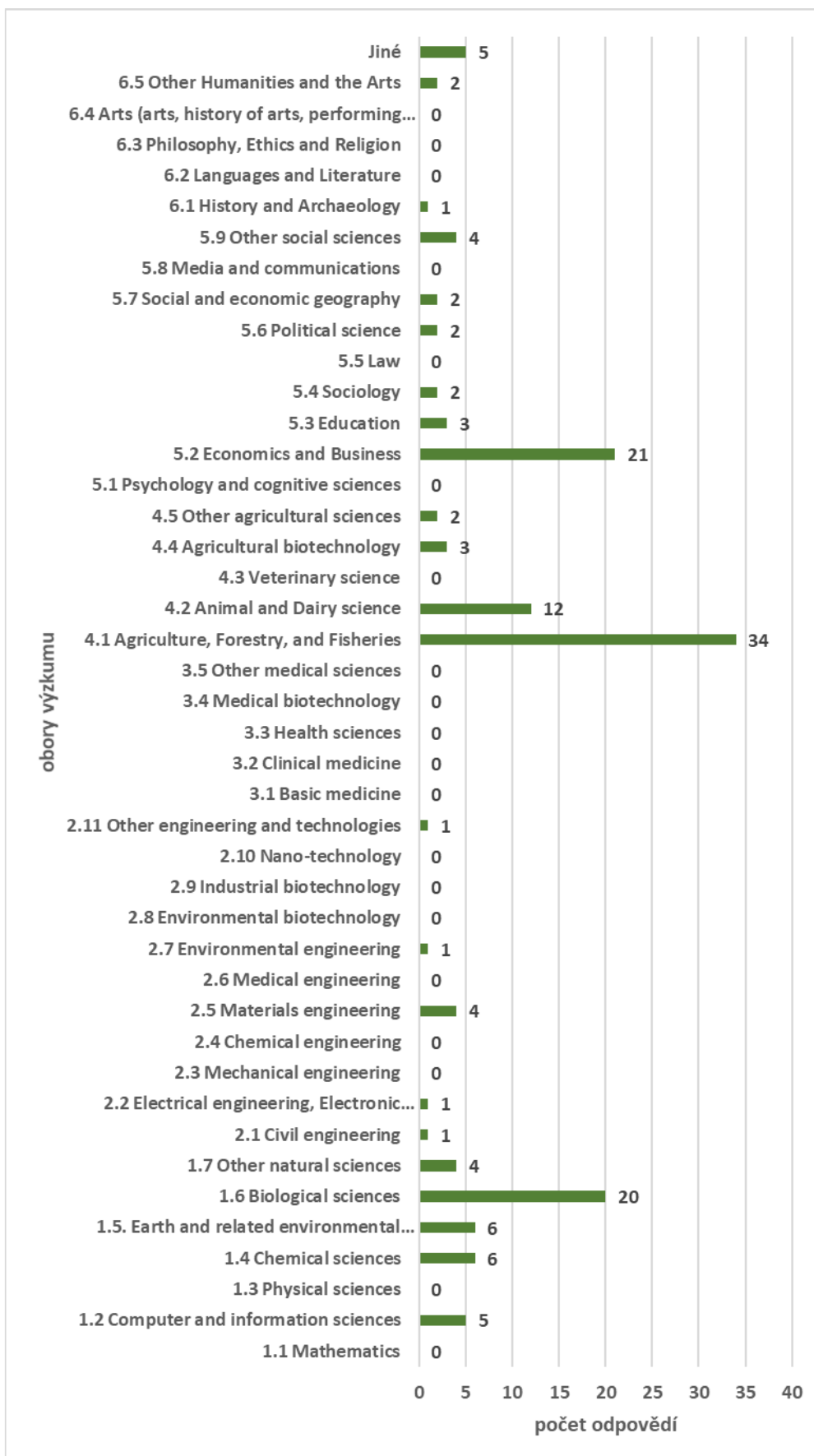
Tabulka č. 2 Porovnání respondentů podle fakult s počtem akademických, vědeckých pracovníků a doktorandů podle fakult.

	AF	LDF	PEF	ZF	FRRMS	ICV
<b>Počet akademických, vědeckých pracovníků a doktorandů (2021)</b>	483	333	268	132	53	9
<b>Počet respondentů</b>	47	37	28	13	13	3
<b>Porovnání účasti na průzkumu k velikosti fakult</b>	9,73 %	11,11 %	10,45 %	9,85 %	<b>24,53 %</b>	<b>33,33 %</b>

### Obor výzkumu

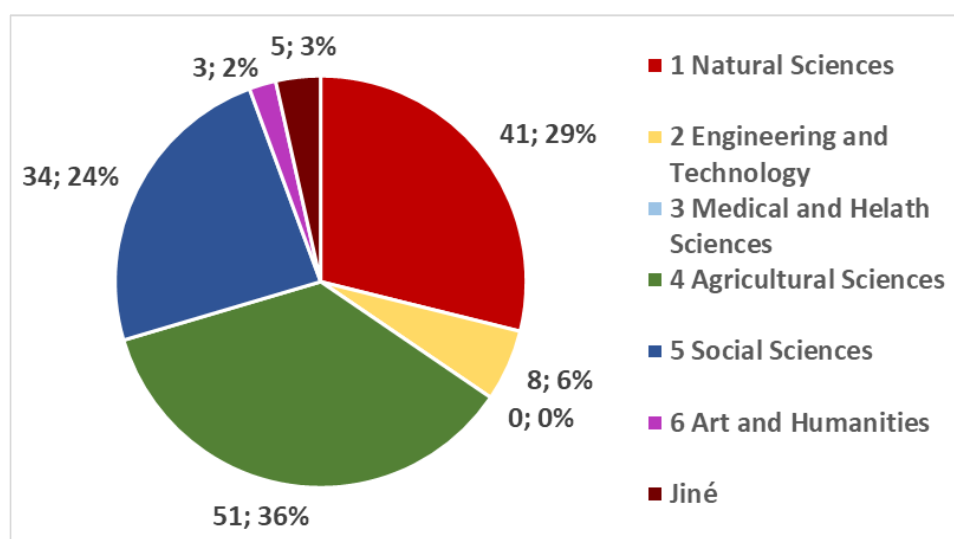
Obory výzkumu byly zvoleny podle oborů FORD (Fields of Research and Development) tak, jak jsou podle Metodiky hodnocení výzkumných organizací vědci zvyklí svoje výsledky odesílat do RIV. Nejvíce respondentů uvedlo obor výzkumu: **4.1 Agriculture, Forestry, and Fisheries (34/142; 23,94 %)**, dále **5.2 Economics and Business (21/142; 14,79 %)** a také **1.6 Biological Sciences (20/142; 14,08 %)**.





Graf č. 4: Obory výzkumu respondentů průzkumu.

Rozdělení oboru výzkumu respondentů podle hlavních OECD kategorií bylo následující: nejvíce se zúčastnili respondenti, kteří provádí výzkum v rámci kategorie 4 Agricultural Sciences (51/142; 35,92 %), 1 Natural Sciences (41/142, 28,87 %) a 5 Social Sciences (34/142; 23,94 %).



Graf č. 5: Rozdělení oboru výzkumu respondentů podle hlavních OECD kategorií.

### Aktuální pracovní pozice

Do průzkumu se nejvíce zapojili **akademičtí pracovníci (110/142; 77,46 %)** a dále **Ph.D. studenti (24/142; 16,90 %)**. Minoritně byli zastoupeni respondenti z kategorie odborný pracovník ve výzkumu (4/142; 2,81 %) a technicko-hospodářští pracovníci (4/142; 2,81 %).

Porovnání aktuálního počtu akademických, vědeckých pracovníků a doktorandů univerzity s počtem respondentů ukazuje, že malé zastoupení respondentů je také u Ph.D. studentů.

	Akademičtí pracovníci	Vědečtí a odborní pracovníci	Ph.D. studenti
<b>FTE</b>	556	199	524
<b>Respondenti</b>	110	4	24
<b>Podíl</b>	<b>19,78 %</b>	<b>2,01 %</b>	<b>4,58 %</b>

Tabulka č. 3: Podíl respondentů a pracovníků dle pracovní pozice.

## Diskuze k oblasti 1. Údaje o respondentovi

### Respondenti dle fakult

Nejvíce respondentů se zúčastnilo z Agronomické fakulty, dále z Lesnické a dřevařské fakulty a z Provozně ekonomické fakulty (graf č. 3). Pro tyto tři fakulty je téma správy výzkumných dat důležité, proto se dal očekávat vyšší podíl respondentů, zároveň se jedná o největší fakulty. Dle zkušenosti organizátorů právě z těchto fakult přichází nejvíce podnětů a dotazů ke správě výzkumných dat.

Ačkoliv podle absolutních čísel byla účast respondentů z ICV a FRRMS poměrně nízká, jejich podíl vzhledem k celkovému počtu akademických, vědeckých pracovníků a doktorandů byl naopak vyšší (24,53 %) (tabulka č. 2). Názory těchto respondentů mají pro vyhodnocení tohoto dotazníku poměrně velký význam.

## Obory výzkumu

Nejvíce se průzkumu zúčastnili respondenti s obory (graf č. 4): 4.1 Agriculture, Forestry, and Fisheries, 5.2 Economics and Business a také 1.6 Biological Sciences. Respondenti nejsou oborově rovnoměrně zastoupeni, na to je potřeba brát ohled při vyvozování závěrů. Oborové zvyklosti se při správě výzkumných dat liší, zastoupení oborů respondentů tedy přímo ovlivňuje výsledky dotazníku.

V průzkumu byly také získány odpovědi z kategorie 5 Social Sciences, u které se jedná o jiná data výzkumu než v kategoriích 4 Agricultural Sciences a 1 Natural Sciences. Žádný respondent nevedl kategorii 3 Medical and Health Sciences, což může souviset s tím, že vědci na univerzitě působí spíše v kategorii 1 Natural Sciences, kam patří i obor (FORD) 1.4 Chemical Sciences.

## Aktuální pozice

Podle očekávání je velké zastoupení akademických pracovníků (tabulka č. 3). Naopak kvůli nízkému zastoupení vědeckých a odborných pracovníků a Ph.D. studentů nejsou dostatečně zjištěny jejich zvyklosti ve správě výzkumných dat.

## Doporučení k oblasti 1. Údaje o respondentovi

1. Zjistit informace o správě dat od pracovníků a studentů univerzity, kteří nejsou dostatečně zastoupeni v průzkumu. Případné doplnění průzkumu o hlubší analýzu.

## 2. Oblast dotazníku: Práce s výzkumnými daty během výzkumu

### Typ a velikost výzkumných dat

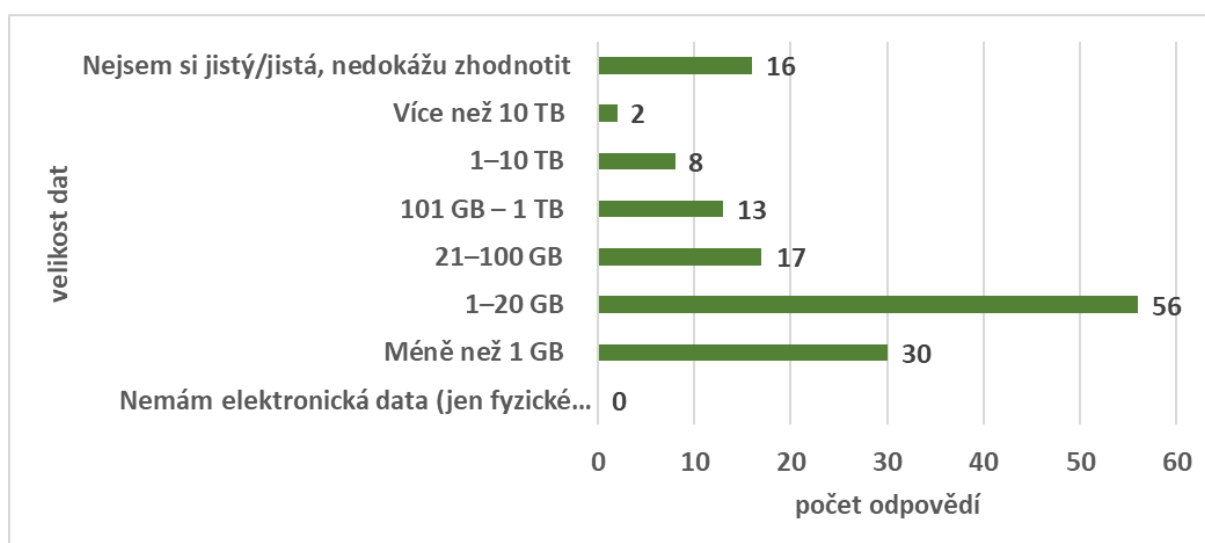
V rámci šetření bylo zjištěno, že respondenti nejčastěji generují tyto data: **text/dokumenty** (134), **tabulky** (132), **obrázky** (119) a **statistická data** (65) (pozn.: bylo možné zadat více odpovědí).

Tabulka č. 4: Typ výzkumných dat.

Typ výzkumných dat	Počet odpovědí
<b>Texty/dokumenty</b> – PDF, TXT, (X)HTML, DOC, DOCX, ODF, ODT, TeX nebo jiné	<b>134</b>
<b>Tabulky</b> – CSV, XLS, XLSX, ODS, SAS, STATA, SPSS, PDF, TXT nebo jiné	<b>132</b>
<b>Obrázky</b> – JPEG, TIFF, JPEG2000, PNG, GIF, BMP, DNG, SVG, PDF nebo jiné	<b>119</b>
<b>Video</b> – MPEG (1/2/4), HEVC (H.265), Flash Video (FLV), QuickTime, WMV, RealMedia, Ogg Video, WebM nebo jiné	<b>32</b>
<b>Audio</b> – LPCM, WAV, MPEG (1/2/4), AAC, DTS, Dolby, FLAC, WMA, RealMedia, Vorbis nebo jiné	<b>15</b>
<b>Strukturovaná data / databáze</b> – XML, JSON, PLAIN TEXT SQL / relační databáze (MS Office Access, MySQL, PostgreSQL, Oracle, MS SQ Server), databáze hostované v cloudu (Amazon DocumentDB, Microsoft Azure Storage Services apod.) nebo jiné	<b>35</b>
<b>Statistická data</b> – DTA, POR, SAS, SAV nebo jiné	<b>65</b>

<b>Specifické typy dat</b> – specifická binární data (výstup speciálního hardwaru/software, např. elektronové mikroskopy) – pro zpracování je třeba specifický/proprietární software, zdrojové kódy softwaru (Python, Ruby, PHP, C, C++, C#, Java, HTML, CSS, JavaScript apod.), zkompileovaný software (binární spustitelné soubory), konfigurační data (INI, CONF, XML, JSON apod.) nebo jiné	<b>26</b>
<b>Jiné</b>	<b>10</b>

Respondenti nejčastěji generují **1-20 GB za rok** (56/142; 39,44 %), ale jsou generována i větší data v řádech TB. Velikost dat se liší mezi jednotlivými obory výzkumu. Někteří vědci si nejsou jistí tím, jak velká data generují (16/142; 11,27 %).



Graf č. 6: Velikost generovaných dat za poslední rok.

### Etické a právní aspekty dat

Průzkum zjišťoval, zda autoři pracují s etickými a právními aspekty při práci s daty. Při této otázce bylo možno zadat více odpovědí. Respondenti dle 78 odpovědí nepracují s chráněnými daty. V případě že respondent pracuje s chráněnými daty, jedná se především o práci s ostatními daty (jiné než osobní údaje) (43), o data třetích stran nebo jsou data předmětem duševního vlastnictví (40) či osobní údaje účastníků výzkumu (32).

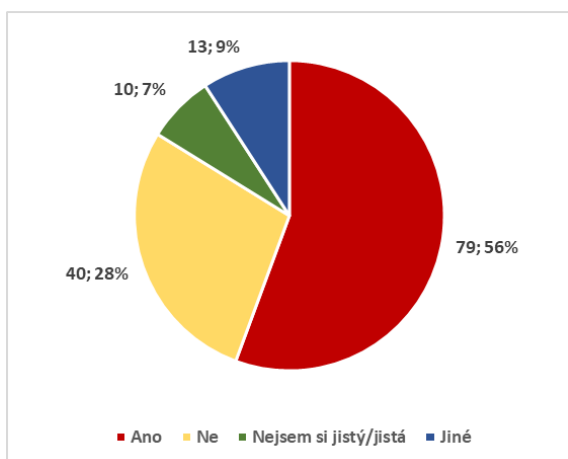
Tabulka č. 5: Etické a právní aspekty práce s daty.

6. Je třeba v rámci vašeho výzkumu řešit specifické otázky týkající se etických či právních aspektů práce s daty? Pracujete s daty, která jsou svou podstatou:	Počet odpovědí
Citlivé údaje (např. lékařské záznamy, údaje o pacientech apod.)	13
Osobní údaje účastníků výzkumu	32
Data z výzkumu s lidskými účastníky (týká se např. respondentů dotazníku i v případě, že nesbíráte osobní údaje)	30
Ostatní data (jiné než osobní údaje)	43
Data související s národní bezpečností (utajované informace podle zákona č. 412/2005 Sb.)	1
Data třetích stran / data jsou předmětem duševního vlastnictví	40
Autorsky chráněná díla	29

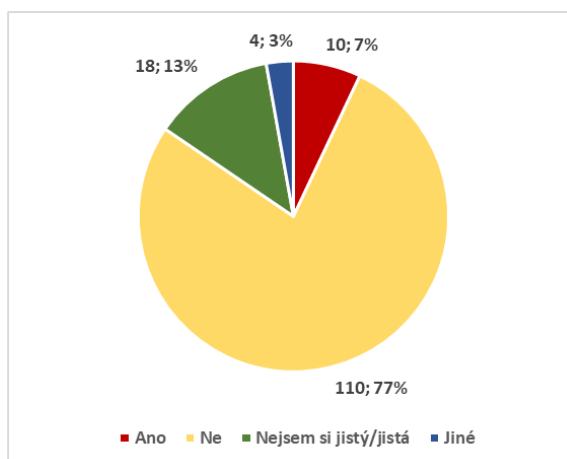
Data, která nepotřebují žádnou speciální ochranu (z možností uvedených výše)	<b>78</b>
Jiné	<b>6</b>

### Dokumentace dat, metadata

Většina vědců vytváří datům **formální strukturovanou dokumentaci** (79/142; 55,63 %), přesto **nepoužívají žádný zavedený oborový standard** (110/142; 77,46 %).



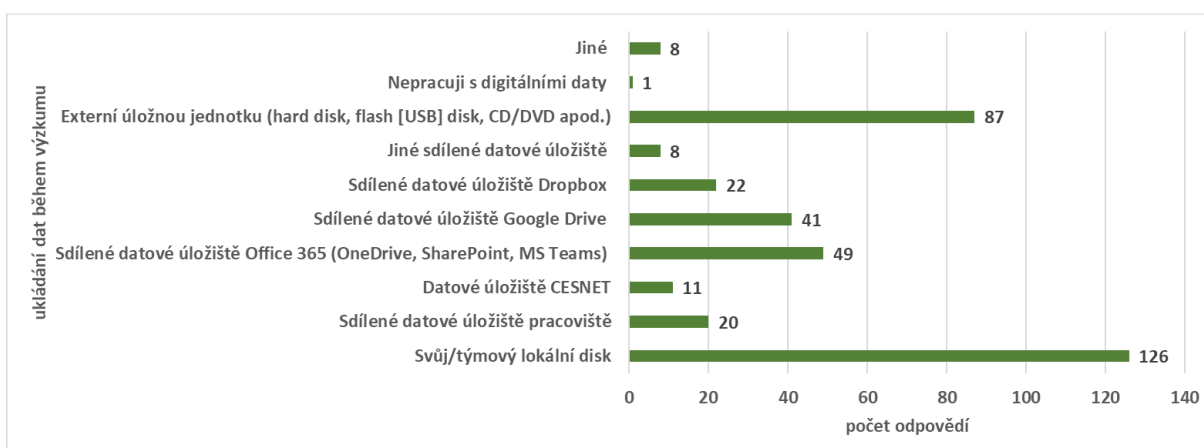
Graf č. 7: Vytváříte ke svým datům také nějaký formální strukturovanou dokumentaci?



Graf č. 8: Používáte k popisu svých dat zavedený oborový metadatový standard?

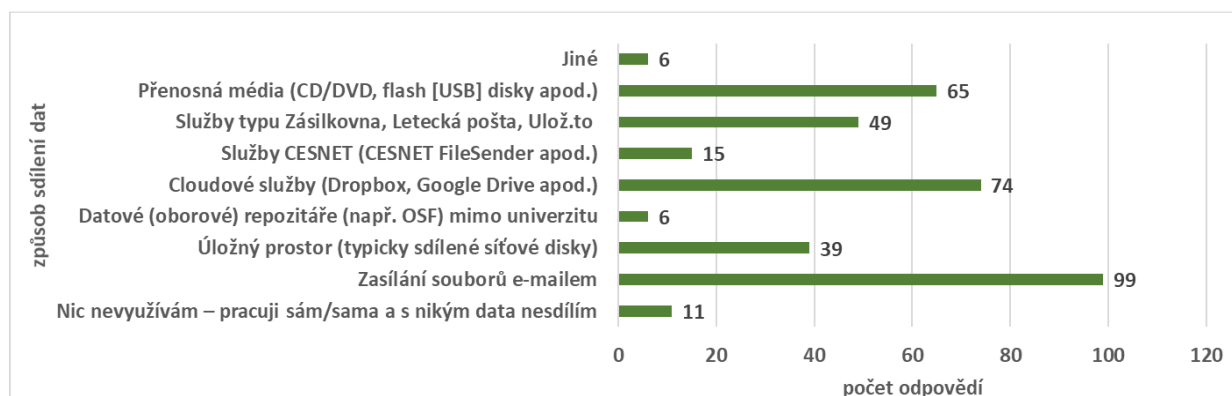
### Ukládání a sdílení dat během výzkumu

U těchto otázek bylo možno si vybrat více odpovědí. Nejčastěji respondenti používají svůj či týmový **lokální disk** (126), dále **externí úložnou jednotku** (87) a **sdílené datové úložiště Office 365** (49).



Graf č. 9: Ukládání dat během výzkumu.

Nejčastěji používaným způsobem **sdílení dat** je **e-mail** (99), dále **cloudové služby** (74) a **přenosná média** (65).



Graf č. 10: Způsob sdílení dat.

### Verzování dat

Většina respondentů **vytváří ručně kopie souborů** s odlišným názvem nebo na jiném místě (102/142; 71,83 %). Celkem 29 respondentů (29/142; 20,42 %) **nevytváří verze datových sad** vůbec, pracuje pouze s aktuální verzí, v případě potřeby data upraví a přepíše. 2 respondenti se k verzování dat nevyjádřili.

### Nástroje pro zpracování dat a vizualizaci

Nejvíce využívaným nástrojem pro výpočty a zpracování dat je Tabulkový procesor (např. Excel) (134 odpovědí), dále Software, který je součástí měřicího přístroje (46), jiné nástroje (45) či programovací jazyk R (30) (pozn.: bylo možné vybrat více odpovědí).

Tabulka č. 6: Používané nástroje pro výpočty a zpracování dat.

12. Jaké používáte nástroje pro výpočty a zpracování dat?	Počet odpovědí
Tabulkový procesor (např. Excel)	<b>134</b>
Power Query	2
Matlab	15
Programovací jazyk R	<b>30</b>
Programovací jazyk Python	11
Software, který je součástí měřicího přístroje	<b>46</b>
Nepoužívám nástroje, počítám ručně (kalkulačka, vlastní poznámky)	11
Jiné	<b>45</b>

Za nástroje pro vizualizace dat nejčastěji respondenti používají Tabulkový kalkulátor (např. Excel) (120 odpovědí), dále Software, který je součástí měřicího přístroje (32), Programovací jazyk R (27). Početná skupina odpovědí (42) uvedla jiné nástroje, například Statistica, ArcGIS (pozn.: bylo možné vybrat více odpovědí).

Tabulka č. 7: Používané nástroje pro vizualizace dat.

13. Jaké nástroje používáte pro vizualizace dat?	Počet odpovědí
Tabulkový kalkulátor (např. Excel)	120
Power BI	2
Tableau	0
Google Data Studio	3
Matlab	16
Programy typu gnuplot, GLE, Matplotlib	3
Programovací jazyk R	27
Programovací jazyk Python	10
Programovací jazyk Python	32
Nevytvářím vizualizace	9
Jiné	42

### Data management plán

Dle výsledků průzkumu většina respondentů Data management plán nezná (134 odpovědí), nebo jej vůbec nevytváří.

### Vybrané slovní komentáře k práci s výzkumnými daty během výzkumu

Ve slovních komentářích odpovědělo 12 respondentů. Respondenti zmínili chybějící úložiště na univerzitě (2/12) a zálohování dat, avšak někteří s povinností ukládat do jednoho úložiště nesouhlasí (1/12). Z otevřených otázek plyne, že by respondenti uvítali pomoc s daty, školení a vzdělávání k práci s daty (2/12) a dále zakoupení licencí pro programy pro práci s daty (1/12). Dále respondenti zmínili komplikované sdílení dat mimo výzkumný tým a univerzitu (1/12) a v jednom případě byly zmíněny služby od CESNET (1/12) pro sdílení dat mimo univerzitu. Pokud má být podporována spolupráce s jinými institucemi, je sdílení dat klíčové. Dále jeden respondent uvedl (1/12), že na univerzitě chybí pracovník pro metodickou podporu v rámci IT a také pro komplexnější analýzu dat.

## Diskuze k oblasti 2. Práce s výzkumnými daty během výzkumu

### Typ dat

Je potřeba vzít úvahu samotnou povahu vytvářených výzkumných dat na univerzitě (tabulka č. 4) a to, že podle oboru výzkumu respondenta se liší i typ generovaných dat a následně způsob správy těchto dat. V dotazníku se respondenti vyjádřili, že nejčastěji generují text/dokumenty, tabulky, obrázky a statistická data, což je rozšířený typ dat.

### Velikost dat

Velikost dat vygenerovaných za rok se liší mezi respondenty (od 1 GB až po velikosti v řádu TB). Z toho důvodu je důležité zjistit konkrétní potřeby vědců a výzkumných týmů ohledně zacházení s daty různých velikostí. Univerzita by měla mít připravenou infrastrukturu jak na menší, tak i na větší objem dat.

## **Etické a právní aspekty dat**

Ačkoliv většina respondentů nepracuje s chráněnými daty (tabulka č. 5), přesto část respondentů tuto zkušenost má. Z odpovědí průzkumu vyplývá potřeba nabídnout vědcům metodickou podporu k nejdůležitějším aspektům ochrany dat, především se jedná o autorsky chráněná díla (patenty, užité vzory), citlivá data z lékařství, údaje v dotaznících a jakékoliv osobní údaje (GDPR). S těmito typy dat je třeba zacházet obezřetně a buď je nelze zveřejnit vůbec (např. autorsky chráněná díla), nebo je potřeba data anonymizovat.

## **Dokumentace dat, metadata**

I když respondenti mají zkušenost s dokumentací dat (graf č. 7), přesto nepoužívají žádný zavedený oborový standard (graf č. 8). To odpovídá zkušenosti organizátorů, že autoři dokumentují svá data spíše podle zvyklostí jednotlivých pracovišť než podle oborových standardů. Pokud ale data nejsou zdokumentována a popsána dle oborových zvyklostí, pravděpodobnost jejich opětovného využití je nižší.

## **Ukládání a sdílení dat během výzkumu**

Téma ukládání dat během výzkumu je klíčové jak pro samotné vědce, tak pro budování infrastruktury univerzity. Vědci by měli mít dostatek nástrojů pro správu dat a možnosti, kam data bezpečně ukládat. Z průzkumu vyplývá (viz graf č. 9), že nejčastěji respondenti používají svůj/týmový lokální disk, dále externí úložnou jednotku anebo sdílené datové úložiště Office 365. Sdílený lokální disk musí být v síti, což sebou nese bezpečnostní riziko. Na druhou stranu je to uživatelsky přívětivé řešení, protože se chová jako „nová složka“. Externí jednotky nejsou optimální řešení, protože jsou náchylné k poškození či se zařízení může ztratit. Sdílená úložiště od Office 365 jsou bezpečná varianta, je zde ale problém přístupu externí osoby a také není vhodný pro objemné datové sady. Zde je potřeba v rámci univerzity nabídnout řešení, kam bezpečně „živá data“ ukládat, a také metodickou podporu, ideálně ve spolupráci s OIT. Respondenti dle průzkumu (graf č. 10) sdílejí data nejčastěji přes e-mail, dále cloudové služby a přenosná média. Sdílení dat pomocí e-mailu není vždy bezpečné řešení, problém může nastat především u citlivých dat a u většího objemu dat. U cloudových služeb může být problém v případě, kdy univerzita nemá smlouvu s těmito službami, univerzita proto nemůže garantovat bezpečnost dat. U přenosných médií hrozí ztráta či poškození nosiče, navíc není možné tímto způsobem jednoduše sdílet data s ostatními.

## **Verzování dat**

Co se týče verzování dat, tak 29 respondentů nevytváří verze datových sad vůbec a pracuje pouze s aktuální verzí, v případě potřeby data upraví a přepíše. Z těchto odpovědí vyplývá, že je potřeba v rámci školení téma verzování vysvětlit.

## **Nástroje pro zpracování dat a vizualizaci**

Nástroje, které vědci využívají ke zpracování a vizualizaci dat, jsou kromě tabulkového procesoru (např. Excel) velmi rozmanité a často respondenti volili možnost Jiné (tabulka č. 6 a 7). Zde se nabízí zjistit potřeby výzkumníků, jaké nástroje používají a jakou další podporu lze nabídnout.



## Data management plan

Většina respondentů Data management plán nezná, nebo jej nevytváří. To odpovídá zkušenostem knihovny z konzultací k tomuto tématu, kdy s Data management plánem pracují většinou jen žadatelé a řešitelé projektů Horizon 2020 a Horizon Europe. Data management plan je ale potřeba vnímat jako proces, tzv. Data management planning, který nastavuje základní správu výzkumných dat.

### Závěry a doporučení k oblasti 2. Práce s výzkumnými daty během výzkumu:

1. Univerzita by měla nabídnout možnosti pro ukládání a sdílení “živých” dat během výzkumu (pro různé formáty a objemy, uživatelsky přívětivé, možnost sdílení i mimo univerzitu, řešení verzování dat).
2. Na univerzitě by mělo vzniknout doporučení pro používání úložišť.<sup>20</sup> Příklad doporučení, které vzniklo na Masarykově univerzitě: <https://it.muni.cz/prehledy/doporuceni-pro-uzivani-ulozist>
3. Vytvoření pracovní pozice metodik správy výzkumných dat (Data steward) v rámci plánovaného Open Science centra, který bude nabízet metodickou podporu v oblasti ukládání dat, verzování atd.
4. Nabídnout podporu v rámci přípravy data management plánu (DMP), nabídka nástrojů pro přípravu DMP.
5. Doporučujeme do budoucna naplánovat možnost technické podpory (IT podpora) pro práci s výzkumnými daty a pro práci s nástroji a infrastrukturou výzkumných dat (úložiště a sdílení dat). Navrhujeme vytvoření tzv. Data helpdesku pro vědce.
6. Provést šetření v rámci dostupných nástrojů pro správu, vizualizaci a statistiku dat. Univerzita by měla nabídnout chybějící nástroje (licence) a školení k práci s těmito nástroji. Toto není v kompetenci Open Science centra, v budoucnu bylo by možné využít Data helpdesku pro vědce, podobně jako na zahraničních univerzitách.
7. Doporučujeme nabídnout pomoc se zpracováním výzkumných dat v rámci výzkumu. Tato podpora by měla být nabídnuta buď v rámci pracoviště vědce, nebo v rámci projektového oddělení/vědy a výzkumu (není v kompetenci Open Science centra).
8. Doporučujeme zvýšit informovanost na univerzitě o právních, etických aspektech a chráněných datech.

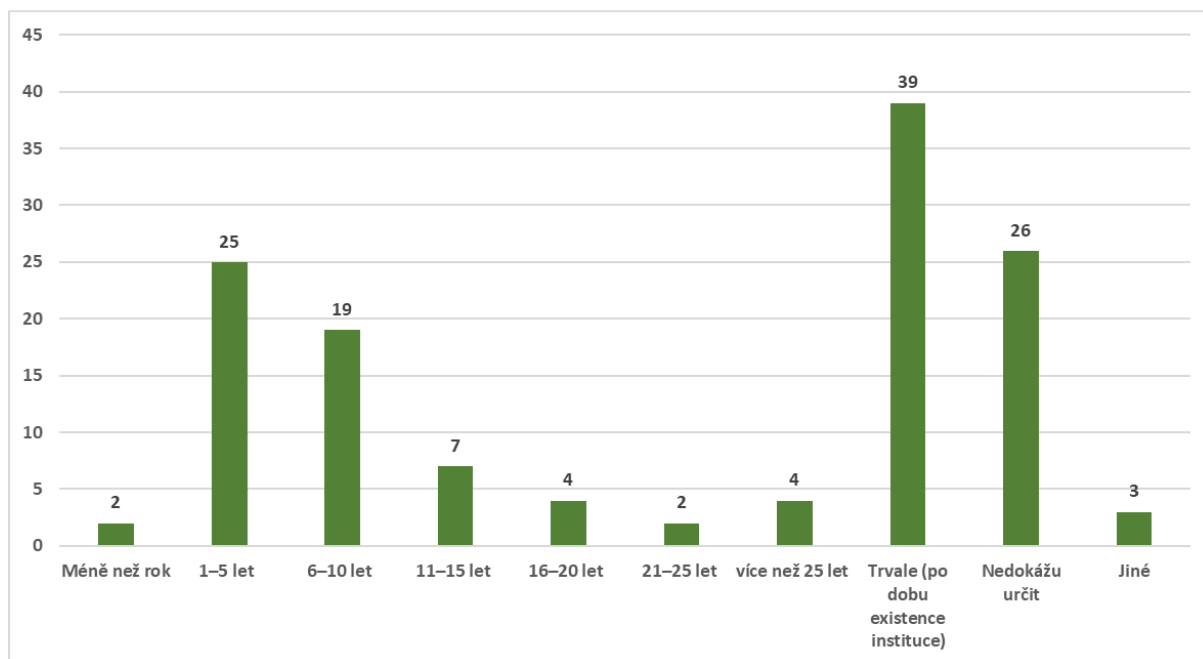
---

<sup>20</sup> Příkladem může být doporučení z Lesnické a dřevařské fakulty: <https://it.ldf.mendelu.cz/?p=128>

### 3. Oblast dotazníku: Dlouhodobé uchování výzkumných dat

#### Časová dostupnost dat

Respondenti mají různé požadavky pro dostupnost a čitelnost dat, nejvíce respondentů uvedlo, že by mělo být uchování trvalé (39/142; 27,46 %). Současně část respondentů nedokáže určit, jaká by měla být dostupnost a čitelnost dat (26/142; 18,31 %). Dále respondenti zvolili relativně krátkou dobu dostupnosti 1-5 let (25/142; 17,61 %) a 11 respondentů (11/142; 7,75 %) se vůbec nevyjádřilo.



Graf č. 11: Požadavky na dostupnost a čitelnost dat.

#### Způsob uchování dat

U otázky k uchování dat bylo možné zadat více odpovědí. Respondenti uvedli, že nejčastěji využívají **externí úložné jednotky** (112 odpovědí), **osobní počítač** (109) a **univerzitní počítač** (63). Pouze 16 respondentů používá univerzitní sdílené síťové jednotky, které jsou zálohované.

Tabulka č. 9: Možnosti uchování dat po skončení projektu.

17. K uchování dat po skončení projektu / fáze výzkumu / výzkumného úkolu využíváte:	Počet odpovědí
Data jsou znehodnocena / smazána	12
Data jsou přepsána	6
Externí úložní jednotku (hard disk, flash [USB] disk, CD/DVD apod.)	<b>112</b>
Osobní počítač (např. lokální pevný disk osobního PC nebo notebooku)	<b>109</b>
Univerzitní počítač (např. lokální pevný disk pracovního PC nebo notebooku)	<b>63</b>
Lokální úložiště (např. na vestavěném pevném disku) univerzitního přístroje/sensory, který generuje data	22
Univerzitní sdílené síťové jednotky (úložný server), jehož obsah je průběžně zálohován.	16

Komerční cloudové služby mimo univerzitu (např. Dropbox, Google Drive / Google Cloud Storage, Amazon Cloud Storage, Microsoft OneDrive / Microsoft Cloud Storage)	46
Prohledatelný digitální archiv (repozitář)	4
Akademické gridové úložiště nebo výpočetní cluster (např. HPC centrum typu CESNET MetaCentrum apod.)	4
Data jsou uchovávána jako fyzické objekty (vytištěné dokumenty v krabicích, chemické vzorky ve skladech apod.)	22
Nejsem si jistý/jistá	2
Jiné	6

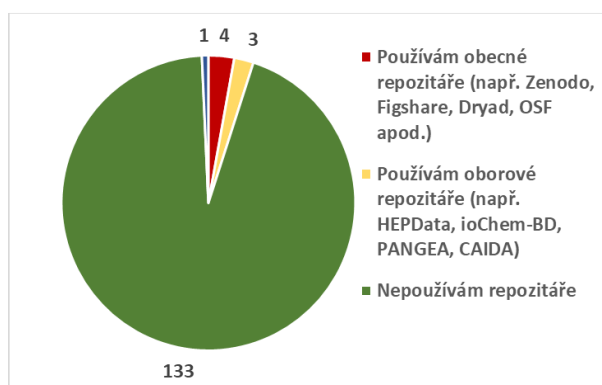
### Metadata

U tématu dokumentace dat **56 respondentů vždy ukládá data společně s veškerou dokumentací (metadata) (56/142; 39,44 %)**, 37 respondentů ukládá s daty pouze část dokumentace (37/142; 26,06 %). Naopak **22 respondentů dokumentaci neukládá (22/142; 15,49 %)** a 11 respondentů neukládá s daty jejich dokumentaci, protože je uložena jinde (11/142; 7,74 %). 1 respondent se nevyjádřil.

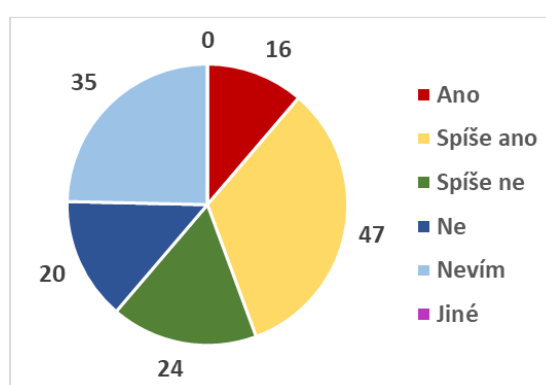
### Využívání datových repozitářů

S tématem uchovávání dat souvisí používání datových repozitářů. **Většina respondentů datový repozitář nepoužívá (133/142; 93,66 %)**. Dále 8 respondentů používá repozitáře PRIDE, OSF, Zenodo, Dryad, ČSDA, Dataverse a Sequence Read Archive GenBank (8/142; 5,63 %).

Oproti tomu **63 respondentů by využilo možnost ukládat data do institucionálního repozitáře MENDELU, kdyby tato možnost existovala (ano a spíše ano 63/142; 44,37 %)**. Poměrně velká část respondentů (spíše ne a ne: 44/142; 30,99 %) by této možností nevyužilo nebo neví, zda by ji využilo (35/142; 24,65 %).



Graf č. 12: Používání repozitáře



Graf č. 13: Využití institucionálního repozitáře

### Uložení dat na univerzitě

Respondenti převážně považují za důležité, aby data byla uložena na univerzitě. Jako nejčastější důvody byly uvedeny: z důvodu trvalého uchování (68 odpovědí), z bezpečnostních důvodů (57) a z technických důvodů (50). Nicméně bylo zatrženo poměrně velké množství odpovědí, že ukládání dat na univerzitě respondenti nepovažují za důležité (43) (pozn.: bylo možné současně zadat více odpovědí).

## **Vybrané slovní komentáře k dlouhodobému uchovávání dat**

Slovní komentář vyplnilo 23 respondentů. Nejčastěji respondenti zmiňovali, že jim chybí místo pro ukládání dat (7/23). Respondenti zdůrazňovali, že data musí být uchována bezpečně a pravidelně zálohovaná (3/23). 1 respondent zmínil, že je potřeba vzdělávání ohledně existujících repozitářů (1/23) a také že by ukládání dat mělo být dobrovolné, spíše služba než povinnost (1/23). Jeden respondent uvedl, že citlivá data ukládá mimo univerzitu na placené služby/servery (1/23).

## **Diskuze k oblasti 3. Dlouhodobé uchovávání výzkumných dat**

### **Časová dostupnost dat**

Nejvíce respondentů uvedlo, že by mělo být uchovávání trvalé. To dává smysl hlavně v mezioborovém výzkumu. Velká část respondentů nedokáže určit, jaká by měla být dostupnost a čitelnost dat, což je dáno nejspíše tím, že dosud nepotřebovali řešit tuto otázku. Z odpovědí vyplývá, že univerzita by měla pokrýt celou škálu požadavků k dostupnosti a čitelnosti uchovávaných dat. Bude třeba identifikovat, která data je potřeba uchovávat krátkodobě a která trvale.

### **Způsob uchovávání dat**

Uchovávání dat po skončení projektu je klíčové téma průzkumu. Jak jsou data uchována po skončení projektu, má vliv na to, zda data jsou do budoucna opět využitelná ať už pro další výzkum, nebo pro doložení či zopakování proběhlého výzkumu. Nejčastěji respondenti využívají externí úložné jednotky, osobní počítač (lokální pevný disk) a univerzitní počítač. U externí úložní jednotky hrozí ztráta dat kvůli krátké životnosti nosiče, připojení externího disku do nejrůznějších zařízení může být nebezpečné (ztráta dat, problém synchronizace zařízení a externí jednotky, možnost zavirování). Navíc není jednoduché sdílet data s ostatními. Sdílená úložiště od Office 365 jsou bezpečná varianta, ale přístup externích uživatelů je problematický.

### **Metadata**

Téma ukládání metadat k výzkumným datům je klíčové pro pochopení samotných dat. Pouze necelých 41 % respondentů ukládá data společně s veškerou dokumentací, to může být zásadní problém pro pozdější správné pochopení nebo opětovné použití dat a pro ověření výsledků výzkumu.

### **Využívání datových repozitářů**

Naprostá většina respondentů datový repozitář nepoužívá, což je při existenci řady repozitářů překvapivé. Uložení výsledku do repozitáře může být povinností, kterou ukládá poskytovatel projektu, vydavatel nebo je to podmínka recenzního řízení. Na druhou stranu téměř 45 % respondentů uvedla, že by použila datový repozitář na univerzitě, pokud by existoval. Do budoucna bude třeba vyřešit, kde bude univerzita výzkumná data dlouhodobě uchovávat, zda vytvoří vlastní datový repozitář. Vedení univerzity by mělo určit, do jaké míry se univerzita zapojí do připravované národní datové infrastruktury a do EOSC a jakým způsobem se zapojí Odbor Informačních technologií.

## Uložení dat na univerzitě

Většina respondentů považuje uložení dat za důležité z různých důvodů, zároveň ale respondenti ve 43 případech zvolili odpověď, že ukládání dat není na univerzitě důležité. Z těchto odpovědí vyplývá, že uložení dat na univerzitě by nemělo být povinné, ale spíše doporučené.

## Závěry a doporučení k oblasti 3. Dlouhodobé uchování výzkumných dat

1. Zajištění bezpečného, trvalého a jednoduchého uchování dat po skončení projektu, které zahrnuje i dokumentaci dat (metadata) a přidělení trvalého identifikátoru DOI. Uchování dat by mělo navazovat na aktivity EOSC.
2. Propagace (oborových i institucionálních) datových repozitářů.
3. Zapojení univerzity do EOSC a do národní datové infrastruktury.
4. Školení k dokumentaci dat, k metadatům a ukládání dat.

## 4. Oblast dotazníku: Sdílení výzkumných dat

### Přístup k datům

**K datům dle respondentů má aktuálně nejčastěji přístup vědecký tým** (110 odpovědí), dále pouze **samotný vědec** (26), zaměstnanci pracoviště (15) ale i také veřejnost (14), protože data jsou veřejně dostupná (pozn.: bylo možné zadat více odpovědí).

Tabulka č. 10 Přístup k datům.

24. Může někdo jiný než vy aktuálně přistupovat k datům z dokončených projektů / fází výzkumu / výzkumných úkolů?	Počet odpovědí
Nikdo	26
Můj výzkumný tým a výzkumní partneři	110
Poskytovatel dotace / zákazník / třetí strana	12
Zaměstnanci pracoviště	15
Zaměstnanci univerzity	5
Jiná specifikovaná skupina osob (profesní komunita apod.)	7
Veřejnost (data jsou veřejně dostupná)	14
Nejsem si jistý/jistá	5
Jiné	4

### Poskytnutí dat

Většina respondentů svá data již v minulosti poskytla/sdílela/uložila, buď na individuální žádost kolegy (90 odpovědí), do časopisu spolu s publikací (74), či s dalšími kolegy na bezpečném místě v rámci univerzity (30). Respondenti zvolili v 17 případech možnost, že svá data zatím neposkytli (pozn.: bylo možné zadat více odpovědí).

Tabulka č. 11 Poskytnutí dat.

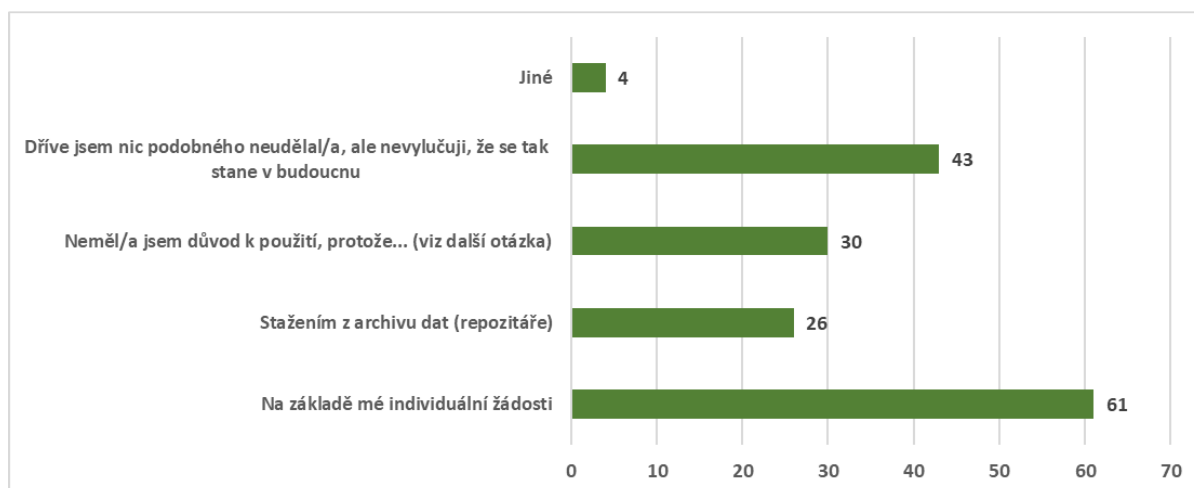
25. Svá výzkumná data jste již dříve poskytli/sdíleli/uložili:	Počet odpovědí
Na individuální žádost kolegy	90
Do časopisu spolu s publikací	74
S některými dalšími kolegy na bezpečném místě v rámci univerzity	30
Na online portálu / repozitáři s chráněným přístupem	20
Na online portálu / repozitáři s otevřeným přístupem	12
Svá data jsem dříve neposkytl/a – nesdílel/a – neuložil/a.	17
Jiné	5

### Používání licencí

Většina respondentů (116/142; 81,69 %) uvedla, že **nepoužívá licencí** při sdílení výzkumných dat. To může souviset s odpovědí na předchozí otázku, podle které mají respondenti především zkušenost se sdílením dat na žádost kolegy, ne veřejně, není tedy licence mnohdy používána. Pokud je použita licence, jedná se většinou o licenci Creative Commons (14/142; 9,86 %).

### Využití dat od jiných vědců

Respondenti odpověděli rozporuplně, zda využili výzkumná data od jiných vědců. Respondenti nejčastěji získali vědecká data na základě individuální žádosti od jiného vědce (61 odpovědí) a dále stažením z archivu dat (26). Naopak ze 43 odpovědí vyplývá, že respondenti výzkumná data od jiných vědců zatím nepoužili nebo respondenti neměli důvod k použití (30) (pozn.: bylo možné zadat více odpovědí).



Graf č. 14: Využití dat od jiných vědců.

Respondenti, kteří neměli potřebu použít data jiných vědců, nejčastěji uvedli jako důvod, že výzkum probíhá na základě vlastních generovaných dat (83 odpovědí) nebo na základě sekundárních dat z veřejné organizace (např. OECD, Eurostat aj.) (34) (pozn.: bylo možné zadat více odpovědí).

## Bariéry pro zpřístupnění dat

Za největší bariéru pro zpřístupňování dat respondenti považují, že data někdo zneužije (55 odpovědí), že je nedostatečný důvod pro zveřejnění dat (54), dále to mohou být obavy z chybné interpretace dat (38) anebo z potenciálních právních či etických problémů (37). Část respondentů je přesvědčena o tom, že příprava výzkumných dat pro ně představuje zátěž (31) (pozn.: bylo možné zadat více odpovědí).

Tabulka č. 12: Bariéry pro zpřístupňování dat.

29. Co považujete za bariéru v případě zpřístupňování vašich výzkumných dat?	Počet odpovědí
Nejsem si jistý, jak na to	26
Má data obsahují citlivé údaje	25
Mám obavy z potenciálních právních či etických problémů	37
Mám obavy, že má data někdo chybně interpretuje	38
Mám obavy, že má data někdo zneužije	55
Příprava dat ke sdílení představuje přílišnou zátěž	31
Nedostatečný důvod pro zveřejňování dat	54
Nedostatečná podpora ze strany univerzity pro přípravu dat ke sdílení	21
Nemám finanční prostředky potřebné pro sdílení dat	14
Jiné	11

## Výhody sdílení dat

Naopak jako výhodu pro sdílení dat vidí vědci možnost využití dat, které nasbíral někdo jiný (65 odpovědí), dále potenciální zvýšení citovanosti související publikace (58) nebo urychlení výzkumného procesu v jejich oboru (56). Poměrně velká skupina vědců vidí výhodu ve sdílení dat v zajištění robustnosti/replikovatelnosti výzkumu (50) (pozn.: bylo možné zadat více odpovědí).

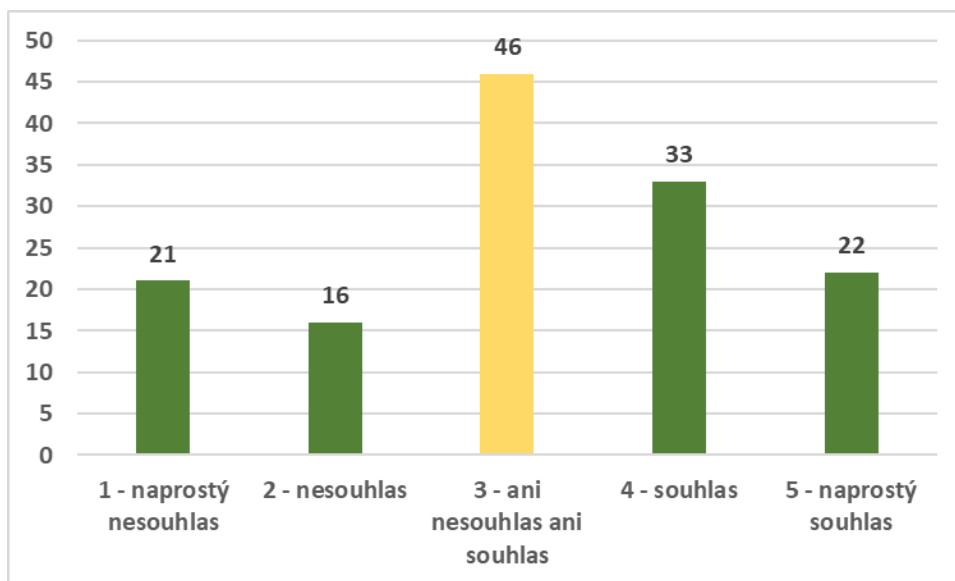
Tabulka č. 13: Výhody sdílení dat.

30. Jaké vidíte výhody sdílení výzkumných dat?	Počet odpovědí
Uspadnění práce při vyřizování žádostí o přístup k mým výzkumným datům	42
Potenciální zvýšení citovanosti související publikace	58
Možnost citovat výzkumná data samostatně, nezávisle na související publikaci	29
Zajištění robustnosti/replikovatelnosti mého výzkumu	50
Urychlení výzkumného procesu v mém oboru	56
Možnost využití dat, která nasbíral někdo jiný	65
Splnění požadavků třetích stran (např. poskytovatele financí či vydavatele)	26
Zvýšení kredibility výzkumu	44
Jiné	15

## Ochota sdílet data v budoucnu

Respondenti odpovídali na otázku, zda jsou v budoucnu ochotní svá data (za určitých podmínek – např. jako anonymizovaná data v agregované podobě apod.) otevřeně sdílet. Respondenti volili odpovědi na škále 1-5 od naprostého nesouhlasu k naprostému souhlasu.

Celkem bylo získáno 138 odpovědí, z nichž převažovala spíše ochota sdílení dat (55/138; 39,86 %). Naopak nesouhlas se sdílením dat byl zaznamenán v menší míře a to 26,81 % (37/138). Poměrně velká skupina respondentů (46/138; 33,33 %) se nedokázala rozhodnout a zvolila střední cestu. Průměrné číslo odpovědí na škále 1-5 bylo 3,14.



Graf č. 15 Ochota sdílet data v budoucnu

#### **Vybrané slovní komentáře k pozitivní zkušenosti se sdílením dat**

Celkem odpovědělo 48 respondentů, pozitivní zkušenost se sdílením dat mělo 8 respondentů (8/48; 16,67 %). Naopak poměrně velká skupina respondentů neměla žádnou pozitivní zkušenost se sdílením výzkumných dat (25/48; 52,08 %). Čtyři respondenti měli pozitivní zkušenosti s využitím dat (4/48; 8,33 %).

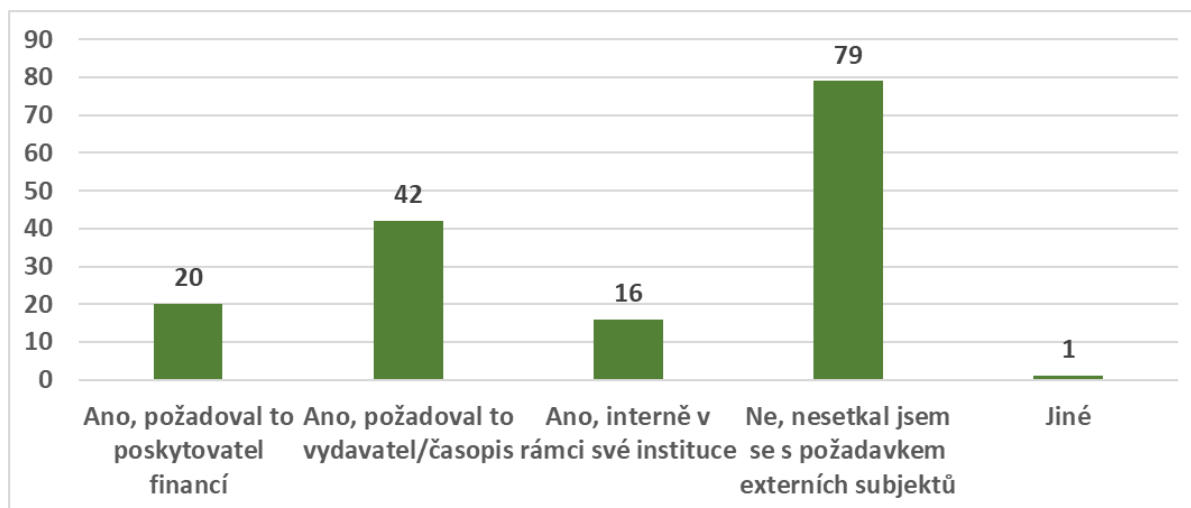
#### **Vybrané slovní komentáře k negativní zkušenosti se sdílením dat**

Celkem odpovědělo 48 respondentů. Respondenti se obávají zneužití dat, 8 respondentů uvedlo své obavy či zkušenosti se zneužitím dat (8/48; 16,67 %). Dvěma respondentům chybí na univerzitě nástroj pro správu dat a považují za vážný problém, že se nesdílí původní naměřená data (2/48; 4,17 %). Zajímavá je zkušenost jednoho respondenta, že vyžádání dat bylo časově náročné, a další respondent měl zkušenost, že vyžádaná data byla nekvalitní. Právě kvalita dat je pro další využití dat klíčová, stejně jako jejich správný a dostatečný popis. Další respondent měl zkušenost s tím, že data poskytl, ale nebyla poté dostatečně citována.

#### **Povinnost sdílení dat**

I když se většina respondentů nesešla s povinností sdílet data (79 odpovědí), někteří respondenti byli vyzváni ke sdílení výzkumných dat vydavatelem (42), poskytovatelem (20) nebo přímo v rámci univerzity (16) (pozn.: bylo možné zadat více odpovědí).





Graf č. 16: Povinnost sdílení dat

## EOSC

Většina respondentů o European Open Science Cloud neslyšela (96/142; 67,60 %). Menší počet respondentů o EOSC slyšel a uvedl, že je pro jeho výzkum důležitý (11/142; 7,75 %). 2 respondenti se nevyjádřili.

## Vybrané slovní komentáře ke sdílení dat

Celkem odpovědělo 10 respondentů. Tři respondenti nepovažují sdílení dat za důležité nebo odrazují od sdílení dat (3/10), dále jeden respondent poukazuje na to, že více než sdílení dat by měla univerzita řešit etickou stránku vědecké (spolu)práce (1/10). Dva respondenti uvedli, že chybí kontaktní osoba pro data na univerzitě anebo by uvítali školení na data (2/10).

## Diskuze k oblasti 4. Sdílení výzkumných dat

### Poskytnutí dat

Z odpovědí vyplývá, že respondenti nejsou zvyklí data poskytovat, a pokud ano, tak především na základě individuální žádosti. Praxe sdílení dat, kdy autor veřejně sdílí data, není na univerzitě běžná.

### Používání licencí

Většina respondentů při sdílení nepoužívá licenci, což souvisí s tím, že respondenti nemají zkušenost se sdílením dat veřejně, ale pouze individuálně na žádost kolegy. Pokud vědci sdílí veřejně výzkumná data, je nutné používat licenci, v opačném případě nejsou data využitelná. Zde se nabízí školení a osvěta k veřejným licencím.

### Využití dat od jiných vědců

Část respondentů neměla důvod využívat data od jiných vědců, je to zřejmě z důvodu, že respondenti v rámci svého výzkumu sami data měří a sbírají nebo používají sekundární data z veřejných organizací. Toto zjištění závisí na oboru výzkumu, v přírodních vědách vznikají data v rámci experimentů, kdežto v humanitních vědách jsou často využita data z veřejných portálů a z databází.

## **Bariéry pro zpřístupnění dat**

Zneužití výzkumných dat jsou legitimní obavy respondentů, o těchto případech by se mělo na univerzitě mluvit a aktivně jim předcházet. Potenciální právní či etické problémy by se mohly vyřešit právní podporou na univerzitě a vědci by se mohli v budoucnu obracet přímo na odborníka. Dále se v odpovědích objevuje vyjádření, že příprava výzkumných dat představuje zátěž a z toho může pramenit neochota vědců výzkumná data spravovat a zpřístupňovat. Řešením by mohla být větší podpora vědcům při správě výzkumných dat.

## **Výhody sdílení dat**

Respondenti si uvědomují také výhody sdílení dat, uvítali by především možnost využití dat od jiných vědců a uvědomují si, že sdílení dat napomáhá k urychlení výzkumu. Z odpovědí vyplývá, že respondenti by raději data využívali, než je poskytovali.

## **Ochota sdílet data v budoucnu**

Převažovala spíše ochota data sdílet, což je pozitivní zjištění vzhledem k Open Science. Celkově ale z odpovědi vyplývá spíše neutrální postoj (průměrné číslo odpovědí na škále 1-5 bylo 3,14). Předpokládáme, že to souvisí s tím, že data nejsou na univerzitě řešena a neexistují pro výzkumníky doporučení, jak postupovat. Také je to poměrně nové téma a respondenti ještě neumí vyhodnotit svůj postoj a potřeby k datům.

## **EOSC**

[EOSC](#) (European Open Science Cloud) není mezi respondenty známý, je tedy úkolem university na EOSC a na připravovanou národní datovou infrastrukturu upozornit a propagovat aktivity a nástroje EOSC vědcům.

## **Závěry a doporučení k oblasti 4. Sdílení výzkumných dat**

1. Nastavení procesů a technické podpory pro sdílení výzkumných dat (Data helpdesk) na univerzitě.
2. Zajištění metodické podpory a vzdělávání k možnostem sdílení dat a používání licencí na univerzitě.
3. Vznik doporučení (best practices) ke správě výzkumných dat. Sdílení dat by nemělo být povinné.
4. Propagace EOSC a národní datové infrastruktury k výzkumným datům.
5. Zajištění podpory v oblasti práva, právní podpora v souvislosti s výzkumnými daty.
6. Vzdělávání k etice výzkumu, k etickým standardům ve vědecké (spolu)práci, práce s etickým kodexem.

## 5. Oblast dotazníku: Podpora správy výzkumných dat

### Podpora při správě dat

Většina respondentů (84/142; 59,15 %) by během svého výzkumu potřebovala odbornou pomoc při práci s výzkumnými daty. Menší část respondentů (56/142; 39,43 %) by žádnou pomoc nepotřebovala.

Při dotazu na konkrétní možnosti podpory respondenti uváděli především obecné informační materiály (62 odpovědí), školení (55) a pomoc se zpracováním dat (55) a podporu v oblasti datových úložišť (44) (pozn.: bylo možné zadat více odpovědí).

Tabulka č. 16: Možnosti podpory při práci s daty.

38. Jaký typ podpory při práci s výzkumnými daty byste rádi využívali?	Počet odpovědí
<b>Obecné informační materiály (návody, checklisty, dobrou praxi apod.)</b>	<b>62</b>
<b>Školení</b>	<b>55</b>
<b>Pomoc se zpracováním dat (HPC, se zpracováním dat pomocí skriptů, konzultaci statistických metod apod.)</b>	<b>55</b>
Specifickou podporu pro správu dat (např. tvorbu data management plánu, výběr vhodného repozitáře, dokumentaci dat apod.)	29
Podporu při práci s citlivými/osobními údaji (např. s jejich uchováváním, anonymizací apod.)	30
Podporu v oblasti etických otázek (např. při práci s lidskými účastníky výzkumu)	23
Právní podporu	31
Podporu v oblasti nástrojů pro tvorbu data management plánů	20
<b>Podporu v oblasti datových úložišť</b>	<b>44</b>
Podporu v oblasti datových laboratoří (Jupyter)	9
Podporu v oblasti nástrojů pro správu a verzování kódu a dat (GitLab)	11
Jiné	9

### Vybrané slovní komentáře k podpoře při práci s výzkumnými daty

Celkem odpovědělo 11 respondentů. 2 respondenti by ocenili školení (2/11), dále se objevila IT podpora (1/11) a pomoc se zálohováním dat (1/11). Jeden respondent nabídl vlastní expertízu pro zpracování výzkumných dat (1/11), z toho vyplývá, že na univerzitě existují vědci s potřebnými zkušenostmi s výzkumnými daty, které je možné oslovit a navázat s nimi spoluprací.

### Diskuze k oblasti 5. Podpora správy výzkumných dat

#### Podpora

Většina respondentů by potřebovala v oblasti správy dat nějakou podporu. Kromě obecných informačních materiálů preferují respondenti školení k výzkumným datům. Dle zkušenosti knihovny ale účast na školeních nebývá vysoká. Neúčast může být způsobená příliš obecnými

tématy školení, která nejsou praktická, nebo je to nedostatečnou propagací školení. Školení by měli vést experti, kteří mají praktickou zkušenost se správou výzkumných dat.

### **Typ podpory**

Respondenti by využili podporu se zpracováním výzkumných dat, toto přání se objevilo i v otevřených odpovědích. V tomto případě je potřeba zajistit potřebnou podporu, ale není úplně jasné, která součást univerzity by ji měla nabízet a v jakém formátu. Nabízí se možnosti zřízení Data helpdesků nebo experta, který pomůže se zpracováním dat. Dále vyplývá z některých odpovědí, že je na univerzitě potřeba metodická a technická podpora v oblasti datových úložišť.

Některá témata respondenti neoznačili jako tolik důležitá, přesto jsou tato témata klíčová pro dobrou správu výzkumných dat: Data management plan, správná dokumentace dat a metadata. U některých odpovědí na otevřené otázky bylo zjištěno, že někteří respondenti nerozuměli tématům citlivých dat a anonymizace dat, právním aspektům (veřejné licence). Také se jeví jako důležité téma výzkumné etiky.

### **Open Science centrum**

V rámci univerzity by mělo vzniknout Open Science centrum, které by mělo vědcům nabízet část požadované podpory, například: obecná školení ke správě dat, výběr repozitáře, licence Creative Commons. Některá témata jsou ale ze své povahy složitější a ty by bylo možné nabízet v rámci jiných součástí univerzity.

### **EOSC**

Další nabídka vzdělávání v oblasti správy výzkumných dat se v budoucnu nabízí v rámci EOSC a připravované národní datové infrastruktury. Vznikne systém vzdělávání jak pro vědce, tak pro Data Stewardy.

## **Závěry a doporučení k oblasti 5. Podpora správy výzkumných dat**

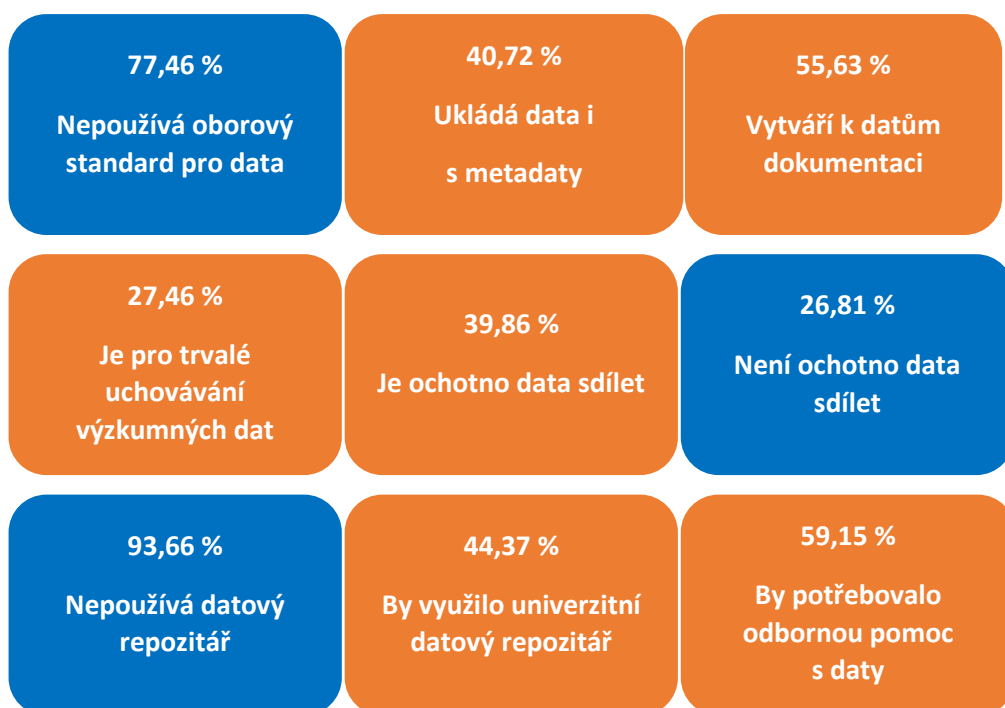
1. Zajištění vzdělávání a školení od expertů, kteří mají zkušenosti se správou výzkumných dat.
2. Vznik doporučení a materiálů pro výzkumná data pro univerzitu a pro jednotlivé fakulty.
3. Vznik Open science centra, napojení na národní datovou infrastrukturu

## Závěrečné doporučení vyplývající z dotazníku, zhodnocení

V budoucnu by bylo vhodné doplnit průzkum o polostrukturované rozhovory s vědci, které by navázaly na témata průzkumu. Je především potřeba podrobně analyzovat, jak akademičtí a vědečtí pracovníci spravují svá data dle oborových zvyklostí. Chybí bližší informace o povaze dat vytvářených na univerzitě (velikost dat, formáty dat, používané oborové standardy při práci s daty atd.).

### Hlavní zjištění dotazníku ke správě výzkumných dat na MENDELU:

Uvedená procenta respondentů



### Nejdůležitější doporučení pro správu výzkumných dat na univerzitě

1. Nastavení správy výzkumných dat na univerzitě, vznik doporučení pro výzkumná data na univerzitě.
2. Univerzita by měla nabídnout možnosti pro ukládání a sdílení (živých) dat během výzkumu (pro různé formáty a objemy, možnost sdílení i mimo univerzitu).
3. Zajištění bezpečného a trvalého uchování dat po skončení projektu, včetně metadat a identifikátoru DOI. Uchování dat by mělo navazovat na aktivity EOSC.
4. Zapojení univerzity do EOSC a do národní datové infrastruktury.
5. Doporučujeme vytvořit Open science centrum a vytvořit pracovní pozici metodik správy výzkumných dat (Data Steward). Cílem je nabízet metodickou podporu a vzdělávání v oblasti správy výzkumných dat.

6. Doporučujeme vytvořit Data helpdesk pro vědce. Cílem je IT technická podpora pro práci s výzkumnými daty a podpora pro práci s nástroji a infrastrukturou (úložiště a sdílení dat).

## Další navrhovaná doporučení

1. Vzdělávání, zapojení expertů, kteří mají zkušenosti se správou dat.
2. Výzkumná data je potřeba řešit komplexně, a to ve spolupráci s Odborem pro podporu tvůrčí činnosti, Odborem Informačních technologií, Výzkumným centrem, Ústavem vědecko-pedagogických informací a služeb apod.
3. Na univerzitě by mělo vzniknout doporučení pro používání úložišť.
4. Provést šetření v rámci dostupných nástrojů pro správu, vizualizaci a statistiku dat. Univerzita by měla nabízet potřebné nástroje a podporu pro práci s nimi.
5. Je potřeba nabídnout podporu se zpracováním výzkumných dat. Tato podpora by měla být nabídnuta buď v rámci pracoviště vědce, nebo Odboru pro podporu tvůrčí činnosti.

## Bibliografie:

GOFAIR. ©2020-2021 Fair Principles. *Gofair.org* [online]. [cit. 2022-06-09]. Dostupné z: <https://www.go-fair.org/fair-principles/>

OPENAIRE. ©2022 How to make your data fair. *Openaire.eu* [online]. [cit. 2022-06-09]. Dostupné z: <https://www.openaire.eu/how-to-make-your-data-fair>

HNÁTKOVÁ, Eva. 2020. Open Science zbožné přání nebo reálná budoucnost vědy. In: *Vedavyzkum.cz* [online]. [cit. 2022-06-09]. Dostupné z: <https://vedavyzkum.cz/nazory-a-komentare/nazory-a-komentare/eva-hnatkova-open-science-zbozne-prani-nebo-realna-budoucnost-vedy>

ELIXIR. 2021. Research Data Management Kit. *Rdmkit.elixir-europe.org* [online]. [cit. 2022-06-09]. Dostupné z: <https://rdmkit.elixir-europe.org/>

UNIVERZITA KARLOVA. ©2022. Podpora. *Centrum pro podporu Open Science*. [online]. [cit. 2022-06-09]. Dostupné z: <https://openscience.cuni.cz/OSCI-112.html>

MASARYKOVA UNIVERZITA. ©2022. Open Science na MU. *Openscience.muni.cz* [online]. [cit. 2022-06-09]. Dostupné z: <https://openscience.muni.cz/>

MASARYKOVA UNIVERZITA. ©2022. O nás. *Ics.muni.cz* [online]. [cit. 2022-06-09]. Dostupné z: <https://www.ics.muni.cz/o-nas/celozivotni-vzdelavani>

RADA PRO VÝZKUM, VÝVOJ A INOVACE. © 2015. Vyhodnocení plnění Akčního plánu. *Vyzkum.cz* [online]. [cit. 2022-07-12]. Dostupné z: <https://www.vyzkum.cz/FrontClanek.aspx?idsekce=876326>

MENDELOVA UNIVERZITA V BRNĚ. © 2002-2022. Úřední deska: Dlouhodobý záměr univerzity. *Mendelu.cz* [online]. [cit. 2022-07-12]. Dostupné z: <https://mendelu.cz/o-univerzite/uredni-deska/uredni-deska-dlouhodoby-zamer-univerzity/?psn=1311>

MASARYKOVA UNIVERZITA, Ústav výpočetní techniky. Přístup vědců ke správě výzkumných dat-průzkum. *Wg-rdm.cz* [online]. [cit. 2022-07-12] Dostupné z: <https://www.wg-rdm.cz/aktivity/pruzkum-sprava-dat-vedci>

MASARYKOVA UNIVERZITA. Doporučení pro užívání úložišť. *It.muni.cz*. [online]. © 2022. [cit. 2022-07-12]. Dostupné z: <https://it.muni.cz/prehledy/doporuceni-pro-uzivani-ulozist>