

Report: **Průzkum uživatelů ČSDA 2022**

Mgr. Michaela Kudrnáčová MSc.



Sociologický ústav AV ČR, v.v.i.

Praha, 2022

Krátce úvodem

Český sociálněvědní datový archiv (ČSDA) Sociologického ústavu AV ČR, v.v.i. odborně zpracovává, dokumentuje a ukládá ze sociálně vědních výzkumných projektů data, propojuje je s relevantními výzkumnými informacemi a kontextem dalších dat a materiálů, a dále také podporuje sekundární analýzu v akademické i vzdělávací sféře. Kromě toho ČSDA pořádá také různé aktivity, jako jsou například semináře, konference nebo workshopy, navíc je zapojen v mnoha projektech různých výzkumných infrastruktur.

Vzhledem k tomu, že se neustále snažíme pracovat na vylepšování našich služeb, je pro nás velmi důležitý názor našich uživatelů. Rozhodli jsme se proto navázat na průzkum ČSDA z roku 2020 a uspořádali podobnou anketu mezi našimi uživateli i v letošním roce. Kromě klasických otázek ohledně využívání našeho datového archivu jsme tentokrát zjišťovali i podrobnosti ohledně FAIR data, open science a uchovávání dat.

Metodologie

Mezi 8. únorem a 31. březnem byl dostupný krátký anonymní online dotazník, v rámci kterého bylo osloveno celkem 3804 uživatelů registrovaných do datového katalogu Nesstar, přičemž 80 z nich bylo vyloučeno z výzkumu z různých důvodů (nedoručitelné pozvánky, změna zaměstnání, na požádání – dlouho archiv nepoužili). Celkový počet respondentů, kteří odpověděli alespoň na jednu otázku, je 281.

Na rozdíl od minulého dotazování byl dotazník tentokrát dostupný jak v české, tak i anglické mutaci. V rámci dotazníku jsme se soustředili pouze na informace relevantní vzhledem k fungování ČSDA, a tak jsme se rozhodli zcela upustit od dotazování se na věk, pohlaví/gender a jiné sociodemografické informace.

Hlavní zjištění

Výsledky průzkumu

- Nejčastěji uživatelé využívají služeb ČSDA v rámci nějakého typu veřejné organizace (84,7 %) jako je veřejná škola, akademická nebo státní instituce.
- Co se týče typu činnosti, převažuje zájem o datový archiv kvůli studiu (32 %).
- Celkem 61,5 % dotazovaných uživatelů využívá výhradně datový archiv ČSDA a žádný jiný.
 - Ti, co využívají i jiné archivy, zmiňovali nejčastěji německý GESIS a slovenský SASD.
- Drtivá většina uživatelů (81,8 %) má vysokoškolské vzdělání.
- Ohledně významu FAIR data mají respondenti sice určitou představu (43,6 %), ale 43,2 % přiznává, že tento pojem nezná a neví, co si pod ním představit.
- Termín Open Science je mezi uživateli, pravděpodobně zejména kvůli výmluvnému názvu, rozšířenější: 46,6 % udává, že tento pojem zná.
- Celkem 56 % respondentů udává, že v rámci svého studia nebo práce produkují nějaká data, ale jen 28,4 % tato data ukládá do nějakého repozitáře nebo datového archivu.
 - Nejčastěji používaným repozitářem je OSF a datovým archivem ČSDA.
- Z těch, co data nikam neukládají, by jich 66,3 % bylo za určitých podmínek ochotných data v budoucnu do repozitáře nebo datového archivu uložit.

Implikace výsledků

- Z relativně velkého množství vyplněných otevřených otázek usuzujeme, že máme podporu uživatelů a že se naši uživatelé mají zájem podílet na rozvoji archivu, sdílet s námi své nápady a zároveň s námi vést diskuzi na téma ukládání dat.
- Zajímavou informací je, že jen cca třetina uživatelů využívá i služeb jiných datových archivů než ČSDA.
- Ze zjištění vyplývá, že je potřeba uživatele intenzivněji edukovat v oblasti FAIR data, ale také Open Science.
- Z průzkumu vyplývá, že je zde velký potenciál získávání dat od našich uživatelů. Mezi překážkami ochoty ukládat data do archivů a repozitářů jsou často neopodstatněné obavy a neznalost existujících procesů. Je třeba lépe představit uživatelům ukládání dat a jeho průběh tak, aby byli potenciální zájemci ujištěni ohledně jednoduchosti tohoto procesu.

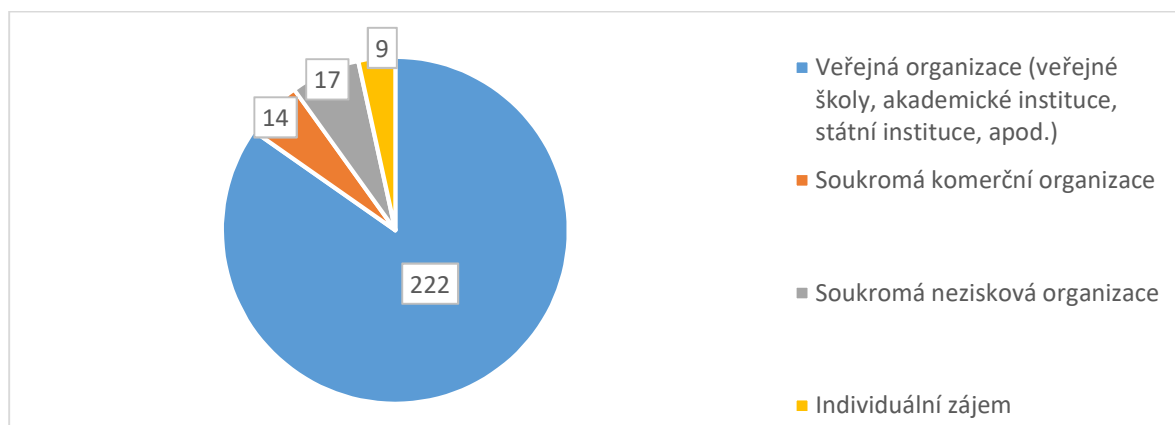
Seznam grafů

Graf 1: Q1. V rámci jakého typu organizace v současnosti využíváte náš archiv?	5
Graf 2: Q2. V rámci jakého typu činnosti v současnosti využíváte náš archiv?	5
Graf 3: Q3. Kromě ČSDA, využíváte i jiné datové archivy k prohlížení/stahování dat?	5
Graf 4: Q4. Jaké je Vaše nejvyšší dosažené vzdělání?	6
Graf 5: Q5. Říká Vám něco označení FAIR data?	7
Graf 6: Q6. Říká Vám něco termín Open Science?	7
Graf 7: Q7. Produkuje v rámci svého studia/práce nějaká data?	7
Graf 8: Q7a. Co se s těmito daty obvykle děje po skončení projektu/fáze výzkumu/výzkumném úkolu?	8
Graf 9: Q7b. Využili jste někdy pro uchování svých výzkumných dat repozitář/datový archiv?	8
Graf 10: Q7c. Jaké platformy obvykle využíváte?	8
Graf 11: Q8. I když jste dosud nevyužil těchto platforem, byl/a byste za určitých podmínek ochotný/á svá data do nějakého repozitáře nebo datového archivu uložit?	9

Seznam tabulek

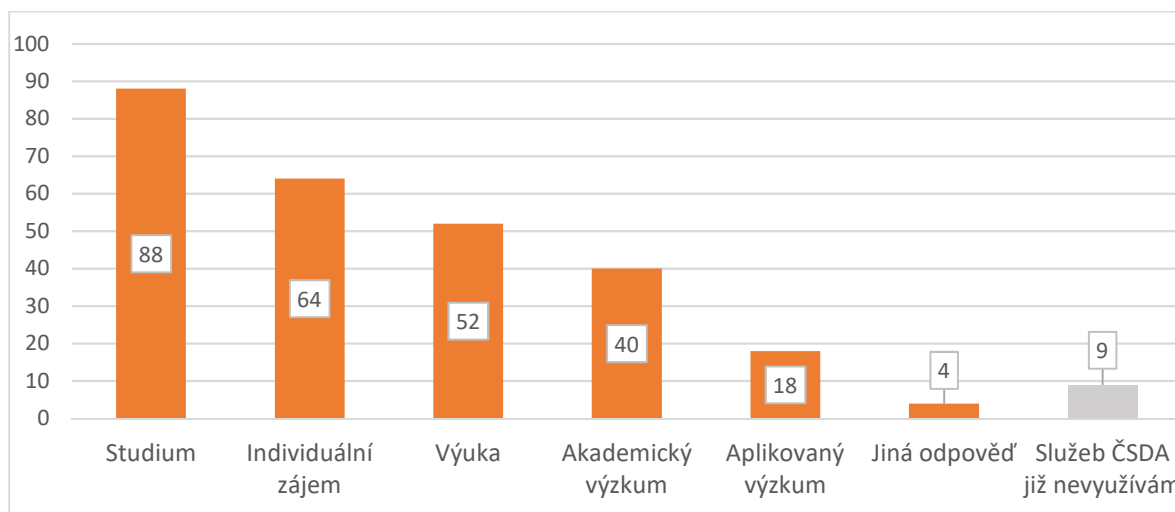
Tabulka 1: Q3a. Které další datové archivy využíváte?	6
Tabulka 2: Q7c1.,Q7c2. Do jakého repozitáře jste data uložili?	9
Tabulka 3: Q7c3. Do jakého datového archivu jste data uložili?	9
Tabulka 4: Q8a. Za jakých podmínek byste byl/a ochotný/á svá data do repozitáře/datového archivu uložit?	9
Tabulka 5: Pokud existuje něco, co byste nám chtěli k tématu uchovávání výzkumných dat sdělit, napište nám, prosím, komentář.	10

Graf 1: Q1. V rámci jakého typu organizace v současnosti využíváte náš archiv?



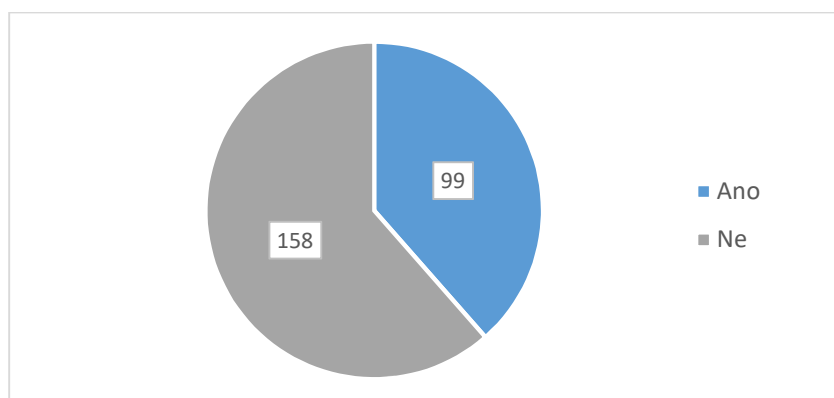
n=262

Graf 2: Q2. V rámci jakého typu činnosti v současnosti využíváte náš archiv?



n=272

Graf 3: Q3. Kromě ČSDA, využíváte i jiné datové archivy k prohlížení/stahování dat?



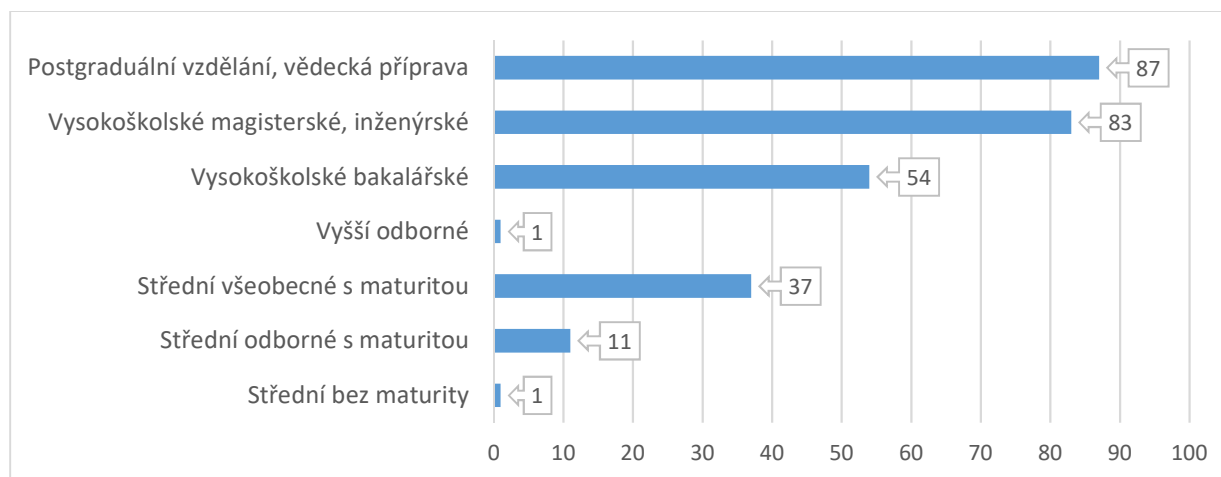
n=257

Tabulka 1: Q3a. Které další datové archivy využíváte?

[Jiné] Které další datové archivy využíváte?	
➤ Německo - GESIS	50
➤ Slovensko - SASD	23
➤ Spojené království - UKDS	11
➤ Rakousko - AUSSDA	9
➤ Nizozemí - DANS	5
➤ Eurostat	5
➤ Norsko - NSD	4
➤ Eurobarometer	4
➤ Itálie - UNIDATA	3
➤ Francie - PROGEDO Research Infrastructure	2
➤ Izrael - ISDC	2
➤ Švédsko - SND	2
➤ Švýcarsko - FORS	2
➤ Estonsko - ESSDA	1
➤ Finsko - FSD	1
➤ Irsko - ISSDA	1
➤ Japonsko - CSRDA	1
➤ Jižní Korea - KSDC	1
➤ Kanada - <odesi>	1
➤ Rumunsko - RODA	1
➤ Srbsko - DSC	1
➤ Španělsko - CIS	1
➤ Ostatní (ESS, ISSP, WVS, speciální projektové, etc.)	13

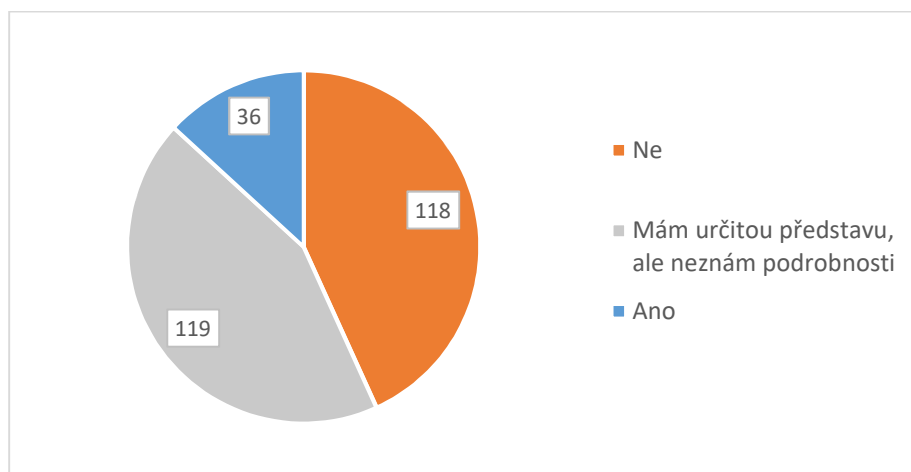
n=144

Graf 4: Q4. Jaké je Vaše nejvyšší dosažené vzdělání?



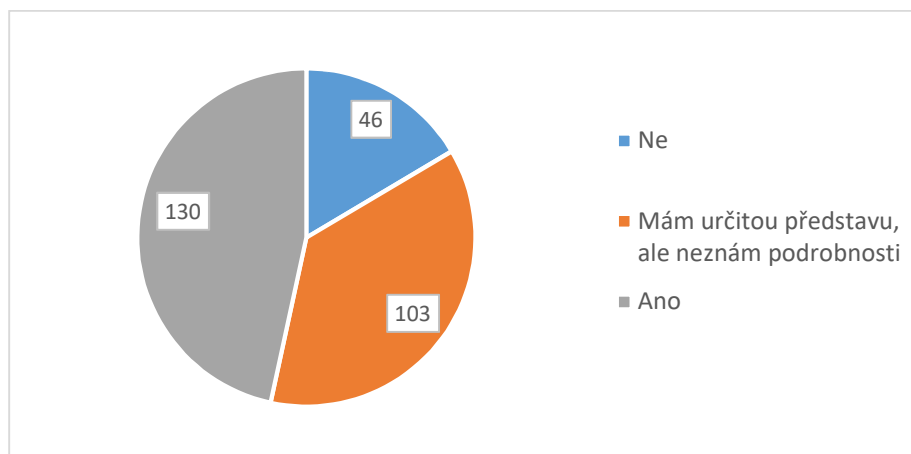
n=274

Graf 5: Q5. Říká Vám něco označení FAIR data?



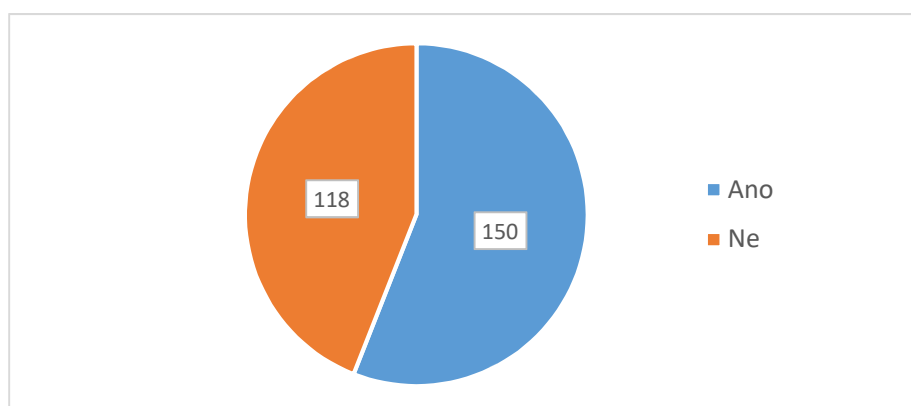
n=273

Graf 6: Q6. Říká Vám něco termín Open Science?



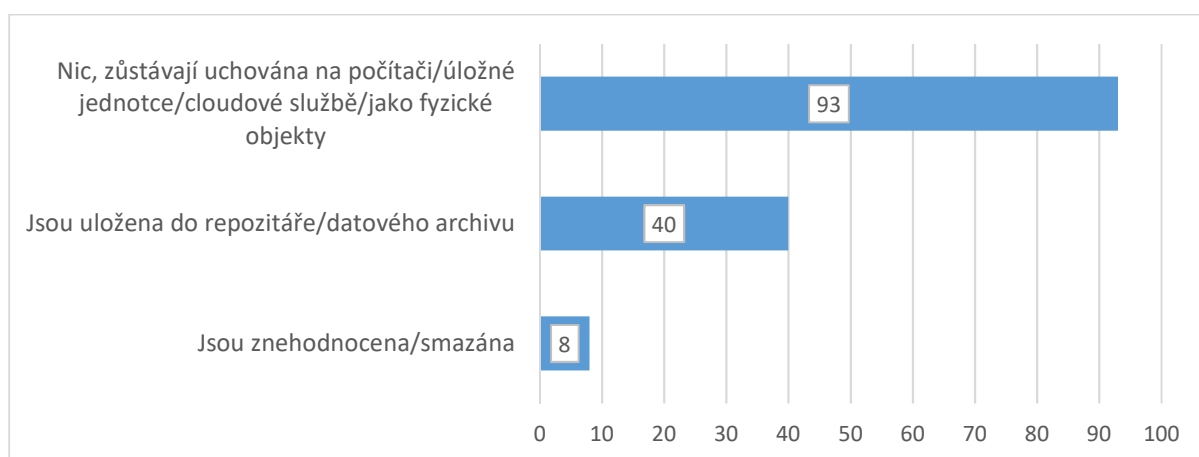
n=279

Graf 7: Q7. Produkujete v rámci svého studia/práce nějaká data?



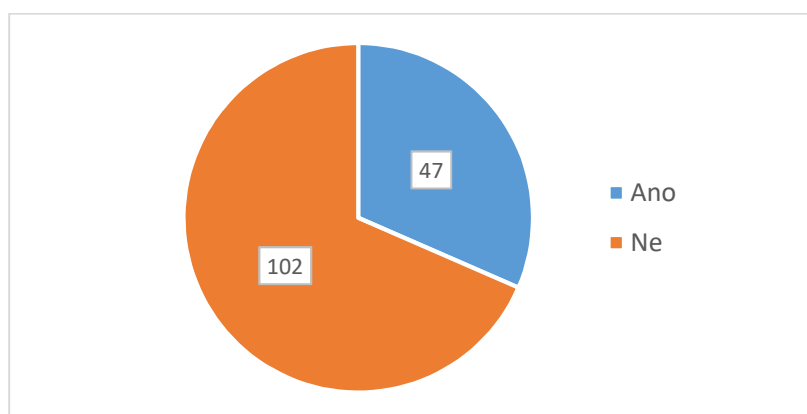
n=268

Graf 8: Q7a. Co se s těmito daty obvykle děje po skončení projektu/fáze výzkumu/výzkumném úkolu?



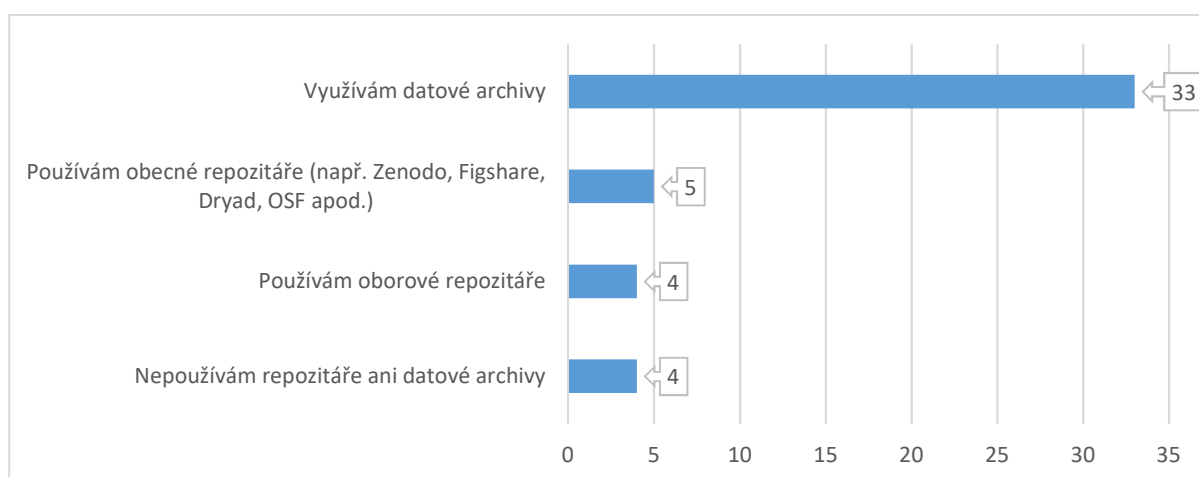
n=141

Graf 9: Q7b. Využili jste někdy pro uchování svých výzkumných dat repozitář/datový archiv?



n=149

Graf 10: Q7c. Jaké platformy obvykle využíváte?



n=46

Tabulka 2: Q7c1.,Q7c2. Do jakého repozitáře jste data uložili?

Do jakého repozitáře jste data uložili?	
➤ OSF	3
➤ Zenodo	1
➤ GitHub	1
➤ Harvard Dataverse	1
➤ Mendeley	1
➤ RepOD	1

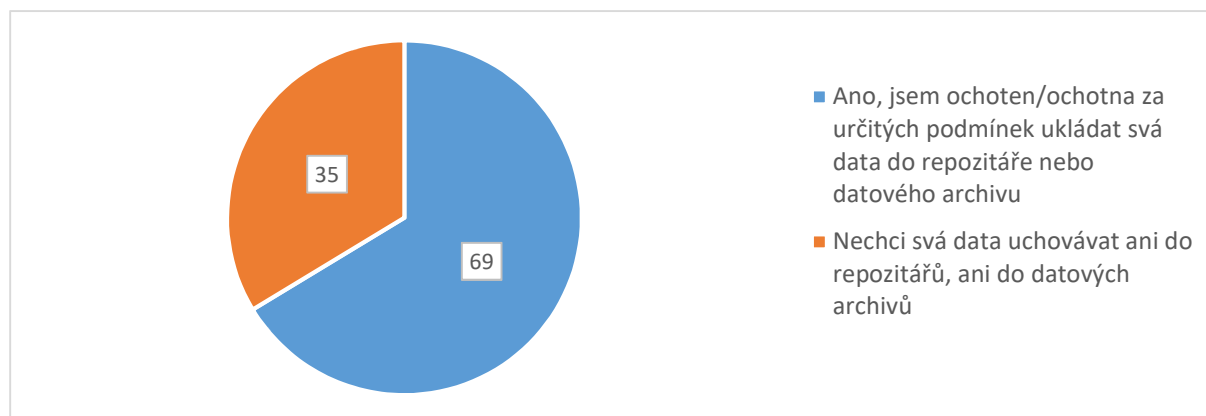
n=8

Tabulka 3: Q7c3. Do jakého datového archivu jste data uložili?

Do jakého datového archivu jste data uložili?	
➤ Česká republika - ČSDA	28
➤ Slovensko – SASD	1
➤ Archiv konkrétního projektu	1
➤ ICPSR	1
➤ Vlastním způsobem	2

n=33

Graf 11: Q8. I když jste dosud nevyužil těchto platforem, byl/a byste za určitých podmínek ochotný/á svá data do nějakého repozitáře nebo datového archivu uložit?



n=104

Tabulka 4: Q8a. Za jakých podmínek byste byl/a ochotný/á svá data do repozitáře/datového archivu uložit?

Otevřené odpovědi
➤ Po jejich základním "vytěžení" v publikacích - Osobně bych uložil dříve/hned, ale organizační a institucionální tlaky většinou požadují jejich využití před zveřejněním
➤ Až budou má data na úrovni, dostatečně reprezentativní a nebudou jen "studentským produktem"
➤ Budou-li to data získaná v projektu financovaném z veřejných peněz. Má aktuální data slouží interně organizaci, která je sbírá.
➤ Budu-li přesvědčen o kvalitě/reprezentativitě mnou poskytovaných dat a v ideálním případě mít i přehled o jejich vynaložení.
➤ Bylo by to pro mě jednoduché a data by byla pro další lidi dostupná zdarma
➤ Jasný přístup, anonymita a etické nakládání s daty. Taktéž možnost kdykoliv poskytnutá data smazat.
➤ Je to dost' složité a napriek množstvu unikátnych dát, ktoré sme získali vo viacerých výskumoch, sme ich do systému nevložili - nemali sme na to kapacity. Tj. spraviť to viac user-friendly a capacity-friendly.
➤ jednoduchost

➤	<i>ještě před výzkumem bych potřebovala vidět, jak jsou data ukládána, abych je sbírala ve formátu, který umí někdo další využít</i>
➤	<i>Když budu vědět, jak na to.</i>
➤	<i>museli by s tímto souhlasit účastníci výzkumů (muselo by být součástí informovaného souhlasu), často uchovávám ale kvalitativní data</i>
➤	<i>muselo by to mít jasný smysl a musel by v některých případech souhlasit komerční klient</i>
➤	<i>Nevím, musel bych o tom vědět více.</i>
➤	<i>nikoli pro komerční využití</i>
➤	<i>Po prostudování způsobu jejich uložení.</i>
➤	<i>Pod heslem</i>
➤	<i>Podle souhlasu respondentu a finanční dostupnosti ukládání</i>
➤	<i>Pokud budu vědět, že dana data jsou pro archiv relevantní a dozvim se, jak je připravit k uchování, není problém je uložit.</i>
➤	<i>Pokud by byla data pořízena s vědomím následného uložení do datového archivu a tato možnost by byla odsouhlasena donátorem výzkumu/sběru dat</i>
➤	<i>Pokud by mi někdo poskytl infrastrukturu a jednoduchý návod, jak to udělat</i>
➤	<i>Pokud by to bylo uživatelsky příjemné a škole by nevadilo sesbíraná data uložit data do datového archivu, rozhodně bych neměla problém. Jako studentka vysoké školy si uvědomuji potřebnost dostupnosti dat.</i>
➤	<i>Pokud by to dávalo smysl, tak ano. U nás sbíráme logy, ty nemá cenu uchovávat dlouhodobě.</i>

Tabulka 5: Pokud existuje něco, co byste nám chtěli k tématu uchovávání výzkumných dat sdělit, napište nám, prosím, komentář.

Otevřené odpovědi	
➤	<i>Že je to skvělá služba a děkuji za vaši práci.</i>
➤	<i>Data, která generujeme, jsou bohužel v drtivé většině případů placena našimi klienty. Vzhledem k obvyklé povaze dat nepředpokládám, že by klienti byli ochotni tato data veřejně sdílet.</i>
➤	<i>ČSDA je pro naše studenty potenciálně důležitý zdroj, zejména kvalitní dotazníky i data z grantových projektů</i>
➤	<i>Určitě bych velmi uvítal: a) méně restriktivní licenci (takhle i použití v médiích je na hraně, natož u vlastního projektu, na který lze třeba přispívat, aby se tak financoval) b) rychlejší publikaci aktuálních dat - aby byla stále aktuální a nebyla to spíše datová archeologie</i>
➤	<i>Naše data jsou z 95 % neveřejná a nelze je sdílet.</i>
➤	<i>Riešim dilemu anonymizácie kvalitatívnych dát a zachovania ich kontextuálnej senzitivity, zatiaľ len v procese bez výsledku. Rada sa od vás poučím.</i>
➤	<i>Možná snažší, jednodušší přístup pomocí klíčových slov k datům v zahraničních výzkumech.</i>
➤	<i>Poslední dvě otázky nesedí mě situaci. Naše firma vytvořila vlastní datový archiv. Pracuji jako výzkumník v komerční soukromé sféře a tyto otázky čili spíše na výzkumníky ve státním, neziskovém a akademickém sektoru.</i>
➤	<i>Sbírali jsme ve škole data jen jednou a styděla bych se je sdílet - ještě to neumíme a sdílet to by byl trapas a hanba, sdílet budu až se to s daty naučím :)</i>